Discretion in Clinical Decision Making: Evidence from Bunching

Claire Boone*

Should healthcare providers strictly adhere to clinical guidelines, or is some discretion beneficial? Using bunching estimation and microdata from 619,907 patients in Chile, I examine discretion in the diagnosis of hypertension (blood pressure$\geq$140/90mmHg). I find bunching in blood pressure at 10% of clinics, where 6 to 62% of patients' test results expected just above the hypertension diagnostic threshold are instead recorded just below it, suggesting providers use their discretion to consider some tests false positives. Leveraging variation in bunching across clinics, I show that affected patients have lower cardiovascular risk, suggesting discretion improves patient classification, partially driven by heuristic thinking.

JEL codes: I10, D01, D81

---

*McGill University, Department of Economics and Department of Equity, Ethics, and Policy. 855 Sherbrooke Street West, Montreal, Quebec, H3A 2T7, claire.boone@mcgill.ca.

I am grateful to Ziad Obermeyer, Jon Kolstad, Paul Gertler, Jason Abaluck, Devin Pope, Nicole Perales, Ulrike Malmendier and three anonymous referees for their helpful comments. I thank the Primary Care Division at Chile's Ministry of Health, and in particular Irma Vargas and Ximena Barros-Rubio. Priyanka Taneja, Madhav Nekkar, and Francesco Crivelli provided excellent research assistance. Funding from NIH National Institute on Aging (NIA) under grant award T32 AG000243 (PI: David Meltzer, MD, PhD). The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIA.

1 Introduction

Clinical practice guidelines aim to standardize medicine using evidence-based recommendations, yet large inefficiencies in care delivery remain. Policymakers argue that improving adherence to guidelines can both reduce costs and improve quality by getting care to the patients who would benefit most [Runnacles et al., 2018]. Critics of guidelines, a group largely made up of physicians, argue that diagnostic rules can be overly prescriptive and ignore the role of expertise, suggesting that non-adherence to guidelines is warranted in some cases [Timmermans and Mauck, 2005].

How much and to which guidelines should health care providers adhere? These questions have important implications for health policy, patient health, and especially for the design of physician payment models that increasingly incentivize guideline-adherent care rather than reimbursing based on the volume of care [CMS, 2016]. Both adherence and non-adherence to guidelines are well-studied, but treatment decisions that purposely depart from guidelines are less well understood [Abaluck et al., 2016, 2020, Chandra and Staiger, 2014, Mullainathan and Obermeyer, 2021]. This kind of discretionary decision making – where providers use their judgment to decide whether to test, treat, or act on the results of a positive test – likely depends on the quality of guidelines themselves and can help inform the extent to which health policy should incentivize strict guideline adherence.

In this paper I study one clinical decision, the diagnosis of hypertension, where departures from guidelines can be identified using bunching estimation. Hypertension, also known as high blood pressure, is estimated to affect over 30% of adults globally [Mills et al., 2020]. It is diagnosed through measuring blood pressure with a cuff, a method that is known to result in many false positives [Baron, 2018].

While the clinical practice guidelines for hypertension are very simple - diagnose if blood pressure is at least 140/90 mmHg - diagnostic mistakes are common because accurately measuring blood pressure can be very challenging [Baron, 2018]. Instantaneous blood pressure, which providers observe in clinic, fluctuates around the patient's long-term or true blood pressure, the latter being a strong predictor of risk for cardiovascular events [Muntner et al., 2019]. Simple actions can change blood pressure within seconds, mostly working to increase it: drinking coffee, sitting with one leg crossed over the other, and for some patients, speaking with a physician [Muntner et al., 2019]. Substantial effort is given to developing systems to improve measurement precision such as averaging multiple measures, making sure the patient is sitting correctly, and waiting a few minutes after the doctor walks in. Still, in a busy primary care clinic where providers spend just a few minutes with each patient, knowing whether a patient testing over 140/90 mmHg is a true positive, and diagnosis with hypertension should occur, or a false positive, and diagnosis should not occur, requires substantial clinical skill.

This noise in blood pressure allows me to identify instances where providers may have used their discretion to diagnose or not diagnose a patient. If a provider believes a patient's test result is a false positive, they may retest until they get a number that aligns with their priors. Or, for some patients, they may take the average of multiple measurements. Or, they might simply record a different number in the medical record than the one actually measured. With enough data, these behaviors are observable as bunching in the distribution of blood pressure under the 140 threshold, which would otherwise be smooth.[1] I quantify this behavior using a bunching estimator and interpret it as a behavioral response to the diagnostic threshold for hypertension. This allows for an investigation into when and how health care providers use discretionary decision-making in one of medicine's many 'grey areas'.

I estimate the magnitude of bunching in blood pressure, my measure of discretionary decision-making, separately at 257 public primary care clinics in Chile, using visits from 619,907 unique patients. Taking advantage of the fact that in Chile's public health care system patients are assigned a primary care clinic based on their home address, I use a double difference approach to estimate the impact of discretion. Specifically, I compare higher vs. lower discretion clinics, and patients recorded as below vs. above the diagnostic threshold.

[1] I expect excess mass below the diagnostic threshold and not above it because blood pressure measured in a medical setting, such as primary care, is well-known to be higher on average than if the same patient's blood pressure was measured elsewhere [Muntner et al., 2019].

I find that approximately 10% of clinics exhibit discretion in the diagnosis of hypertension, and at these clinics, 6% to 62% of patients whose test results were expected to be just above the diagnostic threshold are instead recorded just below it. While it is possible some patients are reassigned from hypertension negative to positive, both provider incentives and the magnitude of bunching below the threshold suggest this behavior should be minimal. These results provide evidence that threshold-based guidelines, which are very common in medicine, can distort provider behavior. I then find that clinical practice guidelines are adhered to similarly at higher and lower discretion clinics. This suggests that the value of blood pressure recorded in the electronic health record does guide decision-making, whether it was manipulated by the provider or not.

Although hypertension itself does not have symptoms, if left untreated, it increases the risk of adverse cardiac events. If discretionary decision-making meant that providers were *incorrectly* reclassifying true positives as false positives we would expect to see more of these hospitalizations among patients at high discretion clinics who were classified below the diagnostic threshold. Instead I find the opposite: a one standard deviation increase in discretion leads to 1.8 percentage point (pp) fewer strokes per 100 patients in the short term (within 3 months), and 6.1pp fewer strokes within a year. Results are similar for heart attacks with 2.8pp and 3.5pp fewer heart attacks per 100 patients, within 3 and 12 months respectively. These results suggest not that discretion itself reduces cardiovascular hospitalization, but that at high discretion clinics, patients who initially tested positive and were reassigned as hypertension negative had lower cardiovascular risk. In other words, providers at high discretion appear to use their discretion to move healthy patients below the threshold while keeping higher-risk patients above it.

How do providers identify low-risk patients who happened to have tested positive? I investigate heuristic thinking as a possible mechanism, finding that patients with characteristics representative of high cardiovascular risk, such as older males, are less likely to be classified as hypertension negative at higher compared to lower discretion clinics. I also observe less bunching in the blood pressure distributions of male patients and those over 65, providing further evidence that providers are less likely to consider their positive tests as false positives. These departures from the clinical practice guidelines, which, in Chile consider blood pressure as the sole determinant of hypertension status, suggest providers incorporate additional information into their decision-making process; behavior that aligns with the representativeness heuristic.

Last, using placebo outcomes, I do not find evidence to suggest that providers use their discretion to make different clinical decisions, sort overall healthier patients under the threshold, or that high discretion clinics provide higher quality care on average, minimizing concerns of unobservable effects of clinic quality. I also do not find bunching in other biomarkers, such as weight and cholesterol, that were measured during the same visit as blood pressure.

This paper contributes to our understanding of how health care providers use expertise and make decisions, sometimes with cognitive shortcuts and rules of thumb. In recent work in this area, Singh [2021] documents heuristic thinking in the delivery room, showing that clinical-decisions are affected by recent adverse events experienced by the provider. Olenski et al. [2020] and Coussens [2018] both find emergency room providers are subject to left-digit bias in age - a heuristic that leads to sub-optimal test decisions. These papers find that heuristic thinking can lead to biased decisions and poor outcomes for some patient-groups. In the setting of hypertension diagnosis, adding more information to the decision process through the use of heuristics appears to be beneficial. This is likely due to the too-simple guideline given the noisy biomarker, which may not be the case for other clinical scenarios studied.

This paper aims to make two additional contributions. First, to our understanding of the allocation of clinical resources. Focusing on the fact that health care providers determine how many and which patients receive medical tests and treatments, this literature uses observational data to examine the volume and cost of care, as well as the quality of clinical decisions. For example, Abaluck et al. [2016] and Chan and Gruber [2020] find that providers' testing decisions are inefficient based on patient risk, while Chandra and Staiger [2014] and Mullainathan and Obermeyer [2021] show heart attack tests and treatment are misallocated in emergency departments compared to an optimal rule. Clinical skill, including the ability to test the 'right' patients is increasingly being recognized as an important and modifiable input to quality health care [Chan et al., 2022, Currie and Zhang, 2025].

It is well known that when incentives or policies change discontinuously at a threshold, behavior can be distorted [Kleven, 2016]. This distortion can occur even when the decision-maker does not benefit directly, but instead distorts their decisions to change outcomes for a beneficiary such as a patient or a student. This paper is closest to Diamond and Persson [2016], who use bunching methods to identify and understand teacher discretion in national test grading. They find that teachers selectively increase some students' scores from below to above the pass threshold. Teachers do this more for students who scored lower than their own average on previous test (i.e. students who 'had a bad test day'), analogous to patients who had a bad blood pressure test day.

Bunching theory and methods have been used to study patients' response to health insurance [Einav et al., 2015, 2017], providers' response to health insurance and incentives [Hernández-Pizarro et al., 2020, Rodríguez-Lesmes and Vera-Hernández, 2021], and hospital wait times [Gruber et al., 2018]. This paper expands this literature by investigating bunching in the context of diagnostic decisions. Given the vast number of medical decisions made using thresholds that categorize continuous biological measurements, there are likely many opportunities to use bunching to understand how and why clinical decisions are made.

2 Setting

2.1 Health and Healthcare in Chile

Chile has a similar burden of disease to other high-income countries: non-communicable diseases account for 82% of years of life lost [PAHO, 2020]. The prevalence of hypertension is approximately 28%, compared to 31% in the United States [Lanas et al., 2020, Ostchega and Nguyen, 2020].

Chile has both public and private healthcare systems. All residents are defaulted in to the public system, where insurance is provided by the federal government and funded primarily through a universal 7% tax on earnings (*Fondo Nacional de Salud*, or FONASA). Approximately 19% of the population opts-out of the public system by purchasing private insurance. Those who remain in the public system are assigned a primary care clinic based on their home address, which lessens concerns about selection into clinics by level of discretion. Primary care is considered to be of relatively high quality and organized, and is obtained either at a clinic, or in the case of rural areas with a single medical center, a hospital [Bossert and Leisewitz, 2016]. Hereafter, I refer to both types of facilities as clinics.

Primary health care received in the public system is low-cost, and copayments are further reduced for chronic diseases, including hypertension, as they are a priority policy area for the federal government [Aguilera et al., 2015]. Medications for chronic diseases, such as anti-hypertensives, are free after a patient is diagnosed.

Supply-side payment and incentives are also standardized nation-wide. Payment from the public insurer to primary care clinics is capitated and determined by the number of registered patients and their risk [JLN, 2017]. Healthcare providers in the public system are employed by the government either on a annual, salaried basis, or a fixed-term contract basis. Some primary care providers also face financial incentives to improve quality of care in key areas. Bonuses can be as large as 20% of pay, however the goals and metrics are not clear and change frequently [JLN, 2017]. Importantly for this project, the bundle of incentives faced by providers does not indicate that there is any reason providers would avoid diagnosing or treating a patient with a (marginally) positive test for hypertension. If anything, they might be more incentivized to diagnose a relatively healthier patient, as this group is then more likely to achieve controlled hypertension.

The Chilean public health care setting therefore offers several advantages: public, primary care clinics are centrally organized by the Ministry of Health meaning data from primary care, hospitals, and pharmacies is available and able to be linked country-wide. This analysis therefore includes a large cross section of patients in terms of location, socioeconomic status, wealth, and health from a population with limited access to care issues. These patients also receive care from clinics that face similar payment structures, incentives, and organization.

2.2 Medical Context

An estimated 1.4 billion people have high blood pressure making it the largest contributor to both all-cause and cardiovascular mortality globally. Despite being a strong cardiovascular risk factor, hypertension lacks symptoms earning it the name 'the silent killer' [WHO, 2013]. For this reason it is common practice to measure patients' blood pressure at *every* primary care clinic visit along with vital signs, as this is the only way to know if blood pressure is high. Chile's clinical practice guidelines define hypertension as blood pressure≥140/90 mmHg [MINSAL, 2010, 2017]. Hypertension can be prevented and treated with low-cost and effective medication and with behavior change including regular exercise and improving diet.

Accurately measuring blood pressure is challenging because instantaneous or short-term blood pressure fluctuates around the patient's long-term or true blood pressure. Blood pressure is influenced in the very short term by stimuli such as caffeine intake, sleep, nervousness, the measurement method, and for some patients, the presence of a health care provider [Muntner et al., 2019]. The influence of a health care provider on blood pressure, also known as 'white coat hypertension', is well documented and typically causes blood pressure to be higher when measured in clinic, due to nervousness or stress, compared to when measured at the patient's home [Muntner et al., 2019]. Providers are alerted to these challenges and to common errors that can lead to inaccurate blood pressure measurement in Chile's clinical practice guidelines (see also table A4).

When making a diagnosis the provider observes instantaneous blood pressure and not true, long-term blood pressure. Substantial effort is devoted to creating training materials (e.g. Stergiou et al. [2020]), technology (e.g. Miao et al. [2020]), and research (e.g. Baron [2018]) to improve the accuracy of blood pressure measured in clinic. Still, it is widely understood that blood pressure is often measured with error. Here, I argue that uncertainty in whether blood pressure measured in clinic accurately reflects the patient's true, underlying health status makes this diagnostic decision one where a providers' clinical skill and use of discretion are crucial.

Noisy measurement contributes to anomalies in recorded blood pressure, including rounding down and left-digit bias; the over-representation of numbers that end in zero, which have long been documented in the medical literature [Alcocer et al., 1997, Wilcox, 1961]. An example of left-digit bias in blood pressure measurement is shown in figure 1, panel B. These numerical anomalies have not been documented in other biomarkers, either those measured at primary care visits (such as pulse), or in laboratories (such as hemoglobin A1c). Figure A4 shows I do not observe bunching or end digit bias in other biomarkers from the same primary care encounters as when patients' blood pressure was recorded. **[Figure 1 here]**

Chile's clinical practice guidelines for hypertension state that patients should be diagnosed if either systolic blood pressure is ≥ 140 mmHg, or diastolic blood pressure is ≥ 90 mmHg [MINSAL, 2010, 2017]. The guidelines also recommend measuring patients' blood pressure twice during a primary care visit, and then averaging the two values. Last, the guidelines recommend patients with elevated blood pressure return for a second visit to confirm the diagnosis. Here, I observe the visit where a new diagnosis occurred (whether the second of two visits or not), and I observe the final blood pressure recorded in the visit (whether an average of two measurements or not). While guidelines state patients should return soon after an elevated blood pressure reading to be diagnosed, my data show this is rare in practice: in the full panel dataset spanning January 2013 to December 2018, among patients not diagnosed at a visit, only 28% return for a second visit in the study period. Among those who do return, average number of days to the next visit is 234 overall (median 143 days), and 240 days among those with blood pressure ≥ 140 (median 130 days), suggesting low guideline adherence on the patient side.

This analysis focuses on systolic blood pressure and not diastolic blood pressure for several reasons: formative research from the Framingham Heart Study concluded that systolic blood pressure was more predictive of cardiovascular risk than diastolic [Kannel et al., 1971]. More recent research has called these findings into question (e.g. Flint et al. [2019]), however the original findings may be sticky. Second, systolic blood pressure may be more salient to clinicians because it is written before or above diastolic blood pressure, i.e. $140/90$ or $\frac{140}{90}$. The salience of systolic blood pressure is reflected in the data, where I observe that diagnosis is more closely aligned with the systolic value rather than the diastolic value. In my patient population, 21% of patients have discordant hypertension, that is, systolic above the diagnostic threshold but diastolic below, or vice versa. Among these patients, the final diagnosis (hypertensive vs. not) aligns with the value of systolic blood pressure in 63% of cases.

3 Data

3.1 Electronic Health Records

I primarily use electronic health records (EHRs) from Chile's public health care system from 2013-2018. These data record routine encounters between providers and patients in primary care and include detailed diagnoses, biomarkers (e.g., blood pressure, blood glucose, weight etc.), demographics (age, sex), comorbidities, and some self-reported health behaviors (e.g., exercise, smoking, alcohol use). Visit data was linked at the patient level to medication prescriptions (including date, drug name, and quantity) and to the universe of hospitalizations (including ICD-10 codes for diagnoses, date, and length of stay) in Chile over the same time period. Hospitalization records are, importantly, from all public and private hospitals nationwide.

Several EHR software providers are used by public clinics, and my data come from the largest provider. Among clinics that use said software, visits in the analysis dataset come from patients in Chile's cardiovascular control program (*Programa de Salud Cardiovascular* or PSCV), which has been operating since 2002.

A patient is considered enrolled in the PSCV once they are diagnosed with one or more diseases of interest: type 2 diabetes, hypertension, dyslipidemia, or have risk factors: a history of cardiovascular disease (CVD), or smoke and are at least 55 years old. PSCV patients continue to seek care at their assigned primary care clinic, and receive additional benefits, namely counseling about how to manage their chronic disease, text-message reminders about upcoming visits, additional tests such as for diabetic foot, and are asked to attend preventative care more frequently than non-PSCV patients [MINSAL, 2017]. For this reason, my EHR data does not represent all patients seen by a provider on a given day, but instead, the interactions providers had with PSCV patients. In figure

A1 I use data from the National Health Survey to show that blood pressure is higher among PSCV eligible patients compared to non-eligible patients, and a smooth distribution, lessening concerns that the bunching results might stem from the EHR sampling frame. See section A.1.1 for more details.

EHR data come from 257 clinics representing 113 municipalities (out of 345 in the country). I link EHR at the municipality level to Chile's National Socioeconomic Characterization Survey (CASEN, 2015 wave), which is representative at the municipality level, to obtain socio-economic characteristics.

To compare the distribution of blood pressure in the EHR to a nationally representative distribution I use Chile's National Health Survey (ENS, 2016-2017 wave), which is also representative of the population at the municipality level.

3.2 Sample

To measure discretion and its impact, my sample includes primary care visits where hypertension could be diagnosed, from patients aged 18-79 at 'large' clinics. Last, I restrict the sample to patients who visited primary care at least 1 year before the end of the study observation period, so that I can measure post-visit hospitalization over a sufficiently long period for all patients.

To do this, I first exclude visits where blood pressure was not recorded (2.4% of visits) and patients under 18 or 80 and above (9.8%). Patients 80 and older are subject to a different diagnostic threshold and goal for controlled blood pressure. Last, I exclude 'small' clinics with too few visits for bunching estimation. I define small as clinics with below the median number of visits (N=1265). This causes 52% of clinics to be excluded, but only 12.5% of patients. This leaves N=619,907 patient-visits at N=257 clinics. See table A1 for further details on sample selection.

Table A3 compares included and excluded clinics in terms of patient and clinic characteristics. Compared to excluded clinics, included clinics are almost two times more likely to be urban, are in wealthier regions, and are in municipalities with more tertiary education; all characteristics correlated with areas that are likely to have large clinics. Included clinics have a higher prevalence of patients with diagnosed type 2 diabetes, perhaps reflecting more healthcare utilization among a wealthier population. The two groups are otherwise similar in terms of patient health.

Importantly, I also exclude visits from patients previously diagnosed with hypertension in order to analyze interactions where a diagnostic decision *could be made*. Among PSCV patients, this leaves individuals with type 2 diabetes (29%; categories are not mutually exclusive as patients can have multiple diagnoses), dyslipidemia (46%), or cardiovascular disease risk factors (63%), who are attending regular care at primary care clinics (table A2). I use each patient's first observed primary care visit to avoid over-representing blood pressure measures that are under 140/90 mmHg, which could impact bunching estimates.

Table 1 describes the study population. The average patient is 58 years old with a body mass index of 30.7, the low end of the obese category. Many are low-income, making less than $320 per month and over a quarter of patients live below the federal poverty line. Compared to the total population of Chile, or even to the population that uses public, primary care, this is a high-risk sample. Due to PSCV inclusion criteria we expect both adverse cardiac events and average blood pressure to be higher in this group than the general public, which may also increase the likelihood of observing bunching in blood pressure. Conversely, it is also likely that providers are more attentive to a high blood pressure reading from this high-risk group compared to the general patient population, which might lead to a lower probability of a behavioral response given a positive test for hypertension. **[Table 1 here]**

4 Empirical Strategy

4.1 Measuring Provider Discretion With Bunching

The bunching approach uses discontinuities in incentives to measure behavior [Kleven, 2016]. Similar to regression discontinuity designs, bunching methods can be applied in settings with a running variable and a threshold that determines treatment. In the case of bunching, the running variable is manipulated (compared to a known counterfactual distribution) and selection with respect to the threshold is present. Under some assumptions, described below, this manipulation can be interpreted as a behavioral response to the threshold.[2]

Here, I study bunching of blood pressure values that are recorded in electronic medical records, and interpret bunching as a behavioral response by healthcare providers to the diagnostic threshold for hypertension. Importantly, what is present here is manipulation of blood pressure that is *recorded by the provider* in the electronic health record, and not precise manipulation of patients' own health.

Using blood pressure as a running variable in which to study bunching offers several advantages. First, because hypertension lacks symptoms and can be deadly, blood pressure measurement is recommended at all primary care visits, alleviating any concerns of selection into testing. Second, as a biological parameter its counterfactual population distribution is known. Third, it is a setting with few frictions. That is, providers measure and record blood pressure themselves and then decide the clinical action taken, if any. Thus, if they wish to, providers can precisely manipulate the measurement and outcome according to their beliefs about the patient.

4.2 Bunching Estimation

Examining raw distributions of blood pressure at the clinic level reveals that clinics vary not only in the amount of bunching, but also in the extent to which providers round blood pressure to numbers ending in zero (figure 1, panel C). Such over-representation of zeros is common in blood pressure measurement and is thought to be driven by use of imprecise measurement tools [Greiver et al., 2019]. This behavior is easily incorporated into bunching estimation to avoid overstating the behavioral response.

To improve the fit of the bunching counterfactual I use a different specification for clinics with a large amount of rounding to zero and for clinics with little rounding to zero. To classify clinics I compare the expected (10%) to observed share of observations that end in zero. I define 'high

[2]The majority of work using the bunching approach is in public finance and studies a behavioral response to taxation (e.g. Chetty et al. [2011], Kleven and Waseem [2013], Saez [2010]).

rounding' clinics as those where end digit zeros make up at least 20% of the clinic's total blood pressure measurements. All other clinics are 'low rounders'. Results are not sensitive to this classification, and are similar if a 15% or 30% threshold is used (see robustness tables A6, A7)

To quantify discretion at each clinic I compare the clinic's observed distribution to the distribution that would exist in the absence of bunching. This counterfactual is recovered by fitting a polynomial regression to the empirical density, excluding observations in a range around the threshold where missing and excess mass is observed. To account for round number bunching I add a set of round number fixed effects. Finally, I predict the counterfactual density for the whole distribution and extrapolate into the bunching region, excluding the contribution of round number fixed effects. For each clinic I separately estimate the following equation:

$$n_i = \sum_{k=0}^{K} \beta_k p_i^k + \sum_{r \in R} \rho_r * \mathbb{1}\{p_i = r\} + \sum_{b=p_l}^{p_u} \gamma_b * \mathbb{1}\{p_i = b\} + \varepsilon_i \tag{1}$$

Where n_i is the number of observations in bin i at a given clinic, k is the order of the polynomial, p_i is the blood pressure midpoint in bin i. The first term models the smooth counterfactual distribution of the running variable using a k-degree polynomial. The second term controls for over-representation at round numbers, where R is a vector of numbers that end in zero, omitting 140 (the threshold) $R = \{80, 90, 100, 110, 120, 130, 150, 160, ..., 200\}$. The third term absorbs all bins in the excluded region of the PDF (area that is affected by bunching), $[p_l, p_u]$, using a set of indicator variables $\mathbb{1}\{p_i = b\}$, each with coefficient γ_b.

I estimate two versions of equation 1: for low rounding clinics data is collapsed into 2-unit blood pressure bins, and for high rounding clinics data is collapsed into 5-unit blood pressure bin. For both, my preferred specification does not exclude bins other than the bin containing the diagnostic threshold at 140, as there is substantial variation across clinics in terms of where *excess* mass is located, but little variation in where *missing* mass is located.[3] In tables A6 and A7 I present alternative specifications, varying the classification of clinics into 'high' vs. 'low' rounders, the

[3]For estimation I use the *bunching* package in R [Mavrokonstantis, 2019]. I use the package's default polynomial of 9. Results are robust to instead using 7-10 but change when a lower order

polynomial used, the bin size, and the exclusion window around the threshold.

The estimate of the counterfactual distribution is the predicted values from (1) after omitting the contribution of the dummies around the threshold, but not omitting the contribution of other round-number dummies. [Kleven and Waseem, 2013]. To quantify the magnitude of discretion, I compare the predicted density, (1), to the observed empirical density. Figure 2 shows two examples of bunching estimation. Importantly, while bunching estimation is typically used to calculate labor supply elasticities, here I simply use bunching to categorize clinics into those with vs. without an observable behavioral response to the diagnostic threshold.

The bunching estimator returns the number of missing, if negative, or excess, if positive, observations at the diagnostic threshold bin divided by the height of the counterfactual. This normalization allows for comparison of magnitudes across clinics with different numbers of visits. The bunching parameter of interest for each clinic, c is:

$$b_c = \frac{\text{N observed patients - N patients in counterfactual}}{\text{N patients in counterfactual}}$$

Each term is measured in the bin containing the diagnostic threshold (140 mmHg). A clinic with $b_c = -0.4$, for example, has 40% of the expected mass at the threshold is missing. Standard errors are obtained by bootstrapping the bunching estimation 100 times per clinic. **[Figure 2 here]**

4.3 Identifying Assumptions for the Bunching Approach

To recover a valid counterfactual for bunching estimation three assumptions must be met [Blomquist et al., 2017]. First, the counterfactual distribution must be bounded and well-behaved. Because blood pressure is a biological measurement this assumption is easily satisfied (see figure A2 for the expected distribution of blood pressure).

---

polynomial is used, which results in excessive smoothing and is not recommended. I also specify that the policy threshold is the minimum of its bin, because the diagnostic threshold is blood pressure$\geq$ 140.

The second assumption is that missing mass is not caused by selection. It is recommended that blood pressure is measured at every primary care visit making selection into testing unlikely. Blood pressure is measured along with vital signs (e.g. pulse, temperature) as a general marker of patient health, and because hypertension lacks symptoms screening is crucial. This fact is reflected in the electronic health records data, where before any restrictions, 97.6% of visits contain a blood pressure measurement (table A1).

Third, manipulation is one-sided and bounded, that is, providers are only rounding blood pressure from above to below the threshold and only for individuals in some range of blood pressure. Because hypertension status is only determined by one threshold, there is no reason for providers to systematically round blood pressure when far from the threshold, providing support for the boundedness assumption.

Here, the direction of rounding is unobservable at the patient-level, however, previous work and the institutional setting provide support for one-sided manipulation. Researchers have compared mean blood pressure in clinics with high left-digit bias (where noisy measurement and thus rounding are more likely) to clinics with low left-digit bias and concluded that blood pressure, if rounded or measured imprecisely, is systematically rounded down [Greiver et al., 2019]. In addition, it is well established that blood pressure is on average higher when measured in clinic compared to at the patient's home [Muntner et al., 2019]. Chile's clinical practice guidelines for hypertension management lists errors that commonly lead to incorrect blood pressure measurement [MINSAL, 2010, 2017]. Of the 8 common errors listed, 7 lead to an overstatement of blood pressure, as large as 20mmHg (see table A4 for more details).

The payment scheme in Chile also suggests rounding down. Healthcare providers have been found to exhibit lower effort when paid a fixed salary compared to other payment models such as fee for service or performance based financing [Das et al., 2016]. Thus, in a salaried setting like Chile's public health care system, providers may avoid diagnosing a marginal patient. While some providers also receive performance bonuses the performance indicators do not include *diagnosis* of hypertension.

For these reasons it is unlikely, but not impossible that providers manipulate blood pressure upwards. If downwards manipulation was masking some upwards manipulation then this is another reason that the behavioral response estimated here using bunching would be a lower bound on the true magnitude of discretion at each primary care clinic.

4.4 Variable Construction

The goal of this analysis is to understand the impact of health care providers' discretion on the quality of hypertension diagnosis decisions. Due to large sample requirements, bunching, which is our measure of discretion, must be measured at the clinic rather than provider level.

After measurement, the magnitude of bunching at the clinic level can be parameterized in several ways. My preferred specification examines the intensive margin: I construct a continuous variable equal to $b_c * -1$ if the clinic's estimated magnitude of bunching was statistically significant at the 5% level, zero otherwise. For easier interpretation, this variable is then normalized across clinics to mean zero, standard deviation one. In robustness checks, I construct four other variables that measure discretion and all lead to similar results (see appendix section A.9 for details). While the second-stage standard errors in my main specification do not account for uncertainty in the estimated bunching magnitudes, the results in Tables A10–A12, which use the FDR-adjusted bunching variable and therefore do not depend on the bunching standard errors, are qualitatively similar.

I study four groups of outcome variables: clinical decisions, hospitalizations, patient characteristics, and placebo outcomes. Clinical decisions include the diagnosis of hypertension and the prescription of anti-hypertensive medication. Hypertension diagnosis appears as a column in the EHR data, and is an indicator equal to 1 for a new diagnosis, 0 otherwise. Prescriptions are retrieved from the national medications database and linked to the EHR by patient ID and date. I construct a variable equal to 1 if a new prescription for anti-hypertension medication was written on the same date as the patient's primary care visit. Importantly, patients do not need to fill the prescription for it to appear in the medications database.

The main health outcomes of interest are hospitalization for stroke, heart attack, or congestive heart failure as these are directly affected by untreated high blood pressure. Because these are measured 3, 6, or 12 months after patient's hypertension screening visit, they are also a measure of cardiovascular risk relative to the patient's current primary care encounter. Placebo health outcomes studied include clinical decisions, biomarkers such as weight and height, all hospitalizations not related to hypertension, and type 2 diabetes management measured by hospitalization for related complications. Placebo hospitalizations exclude any hospitalization for stroke, heart attack, or heart failure. The cause of hospitalization was identified with ICD-10 codes for primary and secondary diagnosis (see table A5 for details). For each diagnosis grouping I then create indicators for if the hospitalization occurred within 3, 6, or 12 months of the hypertension screening visit.

Last, I examine discretion and patient characteristics. I focus on characteristics both observable to the provider and potentially informative of cardiovascular risk: male patient, severe obesity (as measured by body mass index ≥ 35), and age greater than or equal to 55 and 65 years.

4.5 Measuring the Impact of Discretion

To understand how clinical discretion is associated with main and placebo outcomes I use a simple difference-in-differences style framework: I compare outcomes in higher vs. lower discretion clinics and below vs. above the clinical threshold. The coefficient of interest is β_3 in the following regression:

$$Y_{i,c} = \alpha + \beta_1 Discretion_c + \beta_2 \mathbb{1}(BP < 140_i) + \beta_3 (Discretion_c * \mathbb{1}(BP < 140_i)) + X_i + \gamma_i + \lambda_i + \varepsilon_{ic} \tag{2}$$

Where $Discretion_c$ is the normalized excess mass in clinic c, $\mathbb{1}(BP < 140_i)$ is an indicator for if patient i's recorded systolic blood pressure is less than 140 mmHg, 0 otherwise. X_i is a vector of patient characteristics. My preferred specification includes fixed effects for calendar year (γ_i), and quarter (λ_i) of the primary care encounter in all regressions. Regressions for hospitalization outcomes additionally include fixed effects for male and for age in 1 year increments. Standard errors are robust and clustered at the clinic level. All results are robust to the inclusion of no controls, or month indicators, an indicator for if the last digit of recorded systolic blood pressure was zero, or indicators for type of clinic.

This approach takes advantage of the fact that the public healthcare system in Chile does not act as a competitive market. Instead, individuals are assigned to a primary care clinic based on their address of residence. The national insurer, FONASA, keeps a list of all enrollees and their assigned clinic. Because the public health care system does not act as a competitive market, clinics have no explicit incentive to enroll more patients. Therefore, patients are exposed to varying levels of discretion.

Figure 3 provides a test of the assumption that the magnitude of clinics' discretion is uncorrelated with patient, clinic, and municipality characteristics by regressing each characteristic separately on the continuous discretion variable. We see that these characteristics are not significantly predictive of the magnitude of discretion, and that more discretion does not appear to be systematically related to patient health or risk. **[Figure 3 here]**

This approach has some limitations: I must rely on the usual difference-in-differences parallel trends assumption, which here implies that changes in outcomes over systolic blood pressure values are parallel between higher and lower discretion clinics. With a continuous treatment variable (such as discretion), Callaway et al. [2024] show that the standard parallel trends assumption is not enough to rule out the possibility of selection bias. Still, in figure A5, I show that mean unadjusted outcomes in 10-unit bins of systolic blood pressure are roughly parallel between high and low bunching clinics for the ranges where we do not expect outcomes to be affected by discretion (90-120 and 160-200). Means within blood pressure bins for rare outcomes like cardiovascular hospitalizations are somewhat noisy, and due to small cells it is not possible to adjust for key covariates like sex and age here, but these figures help rule out the possibility of very different trends at all levels of blood pressure causing the results I find. At the high end of systolic blood pressure, rates of stroke appear to differ between high and low bunching clinics, but I show in tables A13 -A18 that results are robust to excluding these patients, though power is reduced.

5 Results

5.1 Evidence of Discretion in Blood Pressure Measurement

We begin by examining the magnitude and variation of bunching, the measure of health care providers' propensity to use their discretion in making clinical decisions for hypertension. Figure 1 panel C demonstrates that for some clinics there is clear bunching just below the diagnostic threshold (e.g. clinic B, where blood pressure=139 is highly over-represented), while for other clinics there is little to no bunching (e.g. clinic D).

I quantify the amount of discretion separately for each clinic using equation (1). Among the 257 clinics I study, negative bunching is observed in 140 clinics, and is statistically significant at the 5% level in 24 of those. Negative bunching indicates missing mass at and just above the threshold, suggesting that providers record a lower blood pressure value than they measured. Positive bunching is observed in the remaining 117 clinics, but is statistically significant in only 2. I do not focus on positive bunching because in these two clinics I observe excess mass in the 140-144 bin, and missing mass to the right of that. Because all blood pressure readings at or above 140 indicate the same clinical action, that is, to diagnose hypertension, bunching at 140 is not as clinically meaningful and likely reflects rounding instead of discretion in clinical decision-making. **[Figure 4 here]**

Figure 4 displays the distribution of the magnitude of discretion at the clinic level among clinics with statistically significant negative bunching. The median clinic has a b value of -0.215, with values ranging from -0.06 to -0.622. This means at clinics with blood pressure bunching, 6% to 62.2% of patients who are predicted to have blood pressure at or just above 140 mmHg - and who therefore are eligible for hypertension diagnosis - are instead recorded as having blood pressure under the diagnostic threshold. These results suggest a substantial behavioral response to the diagnostic threshold for hypertension at some primary care clinics, and no behavioral response at others.

The observed pattern of bunching is a combination of several possible behaviors. In the simplest case, the provider measures blood pressure $\geq$140 but instead records a number under 140 in the EHR. Or, the provider may have taken multiple blood pressure measurements for the same patient during the same visit, as suggested in the clinical guidelines. The 'final' blood pressure measurement recorded in the EHR might be an average of multiple readings, or the provider may retest to get to a single number they are satisfied with. Averaging at least two measurements is recommended in Chile's clinical practice guidelines as a method of obtaining a more precise signal of patients' health [MINSAL, 2010, 2017]. Unfortunately, only the final decision is present in the electronic health record. Still, in all outlined scenarios, providers are in theory faced with at least one positive test result, and ultimately record a blood pressure reading below 140 for the patient, indicating they have used their discretion to change the test result.

Importantly, we do not observe the same bunching or rounding behavior in other biomarkers measured at the same encounters in primary care, such as LDL cholesterol and weight (Figure A4). This suggests that the observed bunching in blood pressure measurement is associated with hypertension diagnosis and is not a widespread phenomenon.

5.2 The Impact of Discretion on Clinical Decisions

Given providers' behavioral response to the diagnostic threshold, we next examine the impact of this behavior on clinical decisions. Table 2 displays the impact of discretion on the probability of a new hypertension diagnosis or medication prescription. The coefficient of interest is the interaction between the magnitude of clinics' discretion and the indicator for blood pressure recorded below the diagnostic threshold (Discretion x BP$<$ 140). Importantly, if providers are perfectly following clinical practice guidelines, there will be no diagnoses or prescriptions below systolic blood pressure of 140. And, if providers are following guidelines similarly at high and low discretion clinics, we expect to see a coefficient near zero on the interaction term between discretion and BP$<$ 140.

Indeed, the coefficients on the interaction term for both diagnosis and prescription are each close to zero and not statistically significant, indicating no impact of discretion. That is, providers appear to adhere to clinical practice guidelines similarly at higher and lower discretion clinics. This suggests that the location of a patient's *recorded* blood pressure (whether manipulated or not) with respect to the threshold is meaningful, and likely reflects the provider's belief about the patient's risk. **[Table 2 here]**

5.3 Discretion and Selection by Cardiovascular Risk

When a provider uses their discretion to reassign a positive hypertension test result as negative, they are effectively considering the test a false positive. Untreated high blood pressure raises a patient's risk of adverse cardiac events. Thus, if providers are *incorrectly* reclassifying true positives as false positives we would see a higher rate of hospitalization for cardiac events in the months following the primary care visit in the group affected by discretion.

Instead, I find the opposite. Table 3 (row 1) displays a strong negative association between discretion and the probability of stroke or heart attack per 100 patients. A one standard deviation increase in discretion is associated with a lower probability of stroke hospitalization of 0.02 to 0.06 percentage points, measured within 3 to 12 months respectively (p-value< 0.05). A similar pattern is seen for heart attack, where a one standard deviation increase in discretion is associated with a probability of hospitalization that is 0.028 and 0.035 percentage points lower, within 3 or 12 months, respectively (p-value< 0.1). We observe no effect of discretion on the probability of hospitalization with congestive heart failure. All models are adjusted for patient sex and age, in addition to the controls in equation (1). Results are also robust to the inclusion of controls for type 2 diabetes, dyslipidemia, and reported family history of cardiovascular disease. **[Table 3 here]**

Importantly, these results do not indicate that discretion leads to an *overall* lower number of cardiovascular hospitalizations. Instead, the reduction in the probability of hospitalization among patients classified as hypertension negative at high discretion clinics (table 3, row 1) is offset by an increase in the probability of hospitalization of a similar magnitude among patients classified as hypertension positive at the same clinics (table 3, row 2). These results are unsurprising: as providers at high discretion clinics reassign some *low* cardiovascular risk patients to the hypertension negative group, this leads to a higher concentration of *high* cardiovascular risk patients in the hypertension positive group at high discretion clinics (compared to low discretion clinics), shown as positive coefficients in row 2 of table 3.

Figure A9 plots cardiovascular hospitalizations for stroke or heart attack within one year of the primary care visit, separately by 10 and 20 unit systolic blood pressure bins, and separately by clinics with or without statistically significant bunching at $p < 0.05$. Here, we see that the probability of hospitalization is significantly different only in the blood pressure=150 bin (140-159 mmHg), which corresponds to where we see missing mass in the distribution of blood pressure at high discretion clinics, providing further support that the difference in hospitalizations is due to reassignment of patients on either side of the diagnostic threshold.

The regression results are additionally robust to narrowing the analysis window around the diagnostic threshold: for results conditioning on systolic blood pressure between 90 and 160 mmHg see tables A13, A15, and A17 and between 120 and 160 mmHg see tables A14, A16, and A18.

These results suggest that at clinics where providers are more likely to use their discretion when diagnosing hypertension higher risk patients are more likely to be classified as hypertension positive, and lower risk patients are more likely to be classified as hypertension negative, but there is no net change in the number of hospitalizations.

5.4 Discretion and Placebo Outcomes

The cardiovascular hospitalization results raise several questions, namely, how do providers differentiate between true and false positive tests for hypertension? Is it that providers are sorting overall healthier patients under the threshold, or do they have insight about cardiovascular risk? Could providers at high discretion clinics simply providing an overall higher quality of care?

To answer these questions, I estimate the impact of discretion using equation 2 on several placebo outcomes: clinical tests (whether the patient's blood glucose, cholesterol, weight and height were measured), biomarkers (patient height and weight), and new diagnoses (dyslipidemia and type 2 diabetes) from the primary care encounter, as well as future hospitalizations. I focus on two types of hospitalizations, each measured 3, 6, or 12 months after the primary care encounter:

(1) non-cardiovascular related hospitalization (all hospitalizations excluding stroke, heart attack, and heart failure), a measure of the overall health of patients and (2) the probability of hospitalization for type 2 diabetes causes (excluding stroke, heart attack, and heart failure), among patients who have been previously diagnosed with type 2 diabetes. While an indirect measure of quality of care, if discretion is an indicator of overall clinical skill we would expect patients affected by discretion to also have fewer hospitalizations for their non-hypertension related conditions.

Estimates of the impact of discretion on placebo outcomes are shown in figure 5. All coefficients are precisely estimated null effects, suggesting no relationship between the impact of clinical discretion and these outcomes. The null effects on non-cardiovascular hospitalizations additionally suggest that discretion does not simply lead to overall healthier patients being sorted under the threshold. Similarly, the null effects on diabetes-related hospitalizations, estimated among patients previously diagnosed with diabetes, suggest that providers at high discretion clinics are not more effective at managing diseases other than hypertension, and help to rule out unobservable effects of clinic quality that might relate to discretion. **[Figure 5 here]**

To definitively conclude that discretion leads to a more accurate sorting of patients on either side of the diagnostic threshold in terms of health I would ideally measure effects on patient mortality, however, mortality is unobserved in the analysis dataset. In appendix section A.11 I re-estimate the effect of discretion on cardiovascular hospitalizations among patients who are known to be alive for at least 3, 6, or 12 months after their initial primary care visit. Intuitively, among patients who are alive, hospitalization for stroke or heart attack is one of the most severe outcomes. Table A22 shows similar impacts of discretion on cardiovascular hospitalizations in this sample.

5.5 Discretion and Selection by Patient Characteristics

So far we have seen that providers at high and low discretion clinics are equally likely to follow guidelines for hypertension diagnosis and treatment, and yet, patients with lower cardiovascular risk are more likely to be correctly classified as hypertension negative at high discretion clinics. These two facts can both be true if providers use their discretion to identify false positives and sort these patients below the threshold.

How could providers distinguish between true and false positives given noisy blood pressure readings? When deciding if a patient belongs in the group of false positives providers may, instead of calculating conditional probabilities which is mentally taxing, judge the similarity of the patient to someone who is representative of low cardiovascular risk, i.e. use the representativeness heuristic [Tversky and Kahneman, 1974]. To test for this, I use equation 2 to understand if key patient characteristics are over-represented below the diagnostic threshold at high discretion clinics. Specifically, I examine characteristics that are 1) easily observed by the provider during the patient's visit where hypertension screening occurred, and 2) that would be present in a patient representative high cardiovascular risk. These are male sex, severe obesity, age 55 and older, and age 65 and older.

Table 4 examines the impact of discretion on patient characteristics correlated with cardiovascular risk, interpreted as selection with respect to the threshold. At higher discretion clinics, males are 1.6% less likely to be classified as hypertension negative, as are older patients at a magnitude of 1.0% if the threshold of 55 is used, and 1.5% if 65 is used. **[Table 4 here]**

Consequently, we would expect less bunching in the blood pressure distributions of these observably high-risk groups. To test this, I estimate the magnitude of bunching for male vs. female patients at clinics with statistically significant bunching, as well as for patients aged 55 and older vs. younger than 55, and those aged 65 and older vs. younger than 65. Figure 6 shows that while there is bunching of blood pressure in all groups, the magnitude is smaller among higher-risk groups (males, over 55 and over 65), reinforcing the idea that providers rely on heuristics. However, these differences are not large and the within-group difference is only statistically significant comparing patients under 65 to 65 and older. **[Figure 6 here]**

The age thresholds of 55 and 65 are likely to be salient to providers as these two thresholds appear frequently in Chile's guidelines for managing other cardiovascular diseases [MINSAL, 2017]. I do not find similar results when ages 50, 60, and 70 are used (table A23). Other than for identifying groups with different treatment thresholds ($<$18 and $\geq$80), age is not explicitly included in the clinical practice guidelines for diagnosing hypertension.

These results are consistent with providers using the representative heuristic to differentiate between a true and false positive test for hypertension. However because of the small point estimates, heuristic thinking is likely only part of the provider's decision-making process. As an extension I also investigate the of role private information (section A.13). I find that blood pressure together with patient characteristics associated with high cardiovascular risk explain less than 1% of the variation in future adverse health outcomes, suggesting that providers who are able to make accurate predictions about risk are likely using information that is not available in the EHR, and also not in the guidelines.

5.6 Discretion and Subsequent Visits

The bunching analysis uses an index visit: the first visit a patient is observed where they could be diagnosed. Here, I revisit the longitudinal dataset to understand what happens to patients after they are not diagnosed with hypertension during their index visit. Comparing clinical actions at patients' subsequent visits, I find that 5.5% of patients at low discretion clinics are eventually diagnosed with hypertension, compared to 7.3% at high discretion clinics (Table A27). Many of these patients are diagnosed at their next visit: 47% at high discretion clinics and 50% at low discretion clinics. Last, at low discretion clinics 28.9% of patients are never diagnosed while I observe them in the data, compared to 31.8% at high discretion clinics.

It should be noted that this sample is selected – at high discretion clinics there are more near-positive, undiagnosed patients remaining in the sample. Also, we can only observe subsequent clinical actions for patients who return to a preventative care appointment, a behavior that depends on many factors potentially related to clinical discretion or hypertension risk, so these results are likely biased. Still, the fact a similar share of patients are diagnosed at their next visit reinforces the finding that providers at high discretion clinics were not incorrectly labeling patients as false positives in the short run. In the longer run, we see that more patients at high discretion clinics go on to be diagnosed, possibly reflecting an appropriate delay in diagnosis.

6 Conclusion

This paper documents health care providers' behavioral response to the diagnostic threshold for hypertension. This behavior varies across primary care clinics in Chile, where 6 to 62% of patients' blood pressure measurements estimated to be just above the diagnostic threshold are instead located just below it. These findings suggest providers selectively round some patients' blood pressure down or selectively remeasure blood pressure, turning a positive test result into a negative.

I estimate the impact of discretion on clinical decisions and hospitalizations using a double differences approach. Comparing higher and lower discretion clinics reveals that providers adhere to clinical practice guidelines similarly. Yet, patients at high discretion clinics who are recorded as having a negative test are less likely to be hospitalized with adverse cardiovascular events in the future – the highest risks from uncontrolled high blood pressure. Together these results suggest the providers at high discretion clinics are correctly interpreting some patients' tests results as false positives, which may result in more cost-effective care through avoiding unnecessary diagnoses, or delaying diagnosis and treatment.

The estimated difference in the rate of adverse health outcomes is potentially driven by two mechanisms. Most importantly, at high discretion clinics, false positives are more likely to be correctly labeled as a negative test result. Given no change in the levels of hospitalization, this leaves more true positives above the threshold. Notably, correctly classifying patients in this way would lead to the large negative effects seen, whether the *overall* level of hospitalization at high discretion clinics changed or not.

Second, because over-treatment with prescription drugs can be harmful, not treating false positive patients could also have direct benefits. Recent work has shown there can be harms to over-prescribing: in the United States 12 in 1000 people aged 65 and up were hospitalized for medication harms, and this was largely driven by therapeutic medicines for chronic diseases [Budnitz et al., 2021]. For this reason, and because patients are more likely to adhere to simpler treatment regimes, policymakers have recently turned to treatment simplification and de-prescribing as one method to improve health, especially for cardiovascular diseases [Vrijens et al., 2017].

I also provide evidence suggesting that providers use heuristics to distinguish between true and false positive test results. Males and older patients are less likely to be affected by discretion, and I observe less bunching in their blood pressure distributions.

This paper has several limitations, including the necessity of measuring discretion at the clinic level instead of the provider level. This is done for two reasons: because of the large data requirements for bunching estimation, and because patients are assigned to a public, primary care clinic based on their address, but they are not assigned to a specific provider, causing concerns about selection to providers but not selection into clinics. Still, by aggregating to the clinic level important unobserved heterogeneity in the use of discretion or in clinical skill across providers and within clinic might be masked. However, there are several reasons providers working at the same clinic may behave similarly, such as access to the same technologies within a clinic. Prior work has additionally documented that among physicians, peers' and co-workers' practice style is influenced by others in their immediate network (e.g. Agha and Zeltzer [2019]). Additionally, geographic variations in provider preferences and practice style are well documented [Baicker et al., 2004, Finkelstein et al., 2016]. As such, the measure of discretion here is a combination of clinical skill and practice style at each primary care clinic.

Other limitations include the existence of only the 'reported' blood pressure measurement and not all measurements taken at the encounter, if there were multiple, or information about the provider's thinking, resulting in a somewhat of a black box of clinical decision-making. Also, I do not observe patient mortality, which may be a more objective measure of health than cardiovascular-related hospitalizations. However, because care is low or zero cost in the Chilean setting, concerns about access to care should be limited. Similarly, while hospitalization results suggest providers' discretion is applied to false positive patients, I do not estimate the returns to hypertension treatment for marginally undiagnosed patients or an optimal treatment threshold, making it difficult to determine whether discretion improves or worsens patient outcomes in the long term.

The question of the extent to which providers should adhere to clinical practice guidelines has significant implications for health policy. Recent efforts have focused on realigning physician incentives to prioritize adherence to guidelines and improve patient health outcomes, aiming to move away from incentivizing the quantity of care provided [CMS, 2016]. However, this paper highlights a clinical scenario where non-adherence to guidelines can be advantageous for patients in certain situations. The quality of guidelines, and the tools or biomarkers used to measure patient health, are important to consider in such payment policy.

Many medical decisions exist within ‘grey areas’ characterized by high levels of uncertainty. This uncertainty often arises due to the necessity of making decisions with limited information about patient health. For example, the decision to prescribe antibiotics is usually made before test results return from the lab, and costly test decisions, such as testing for heart attacks, rely heavily on subjective information provided by the patient. In both of these cases, research has shown that prediction algorithms could improve the quality of clinical decisions [Huang and Ullrich, 2021, Mullainathan and Obermeyer, 2021]. A noisy biomarker or too-simple clinical practice guidelines also create uncertainty. My findings suggest that healthcare providers pull from the rich set of information that may not be included in clinical guidelines or the electronic medical record when making decisions. As decision support tools, including algorithms and payment models that prioritize guidelines adherence become increasingly prevalent in medicine, it is crucial to understand the quality of guidelines and study the nuances of whether provider deviations from guidelines ultimately benefit or harm patients.

7 References

Jason Abaluck, Leila Agha, Chris Kabrhel, Ali Raja, and Arjun Venkatesh. The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care. *American Economic Review*, 106(12):3730–3764, December 2016. ISSN 0002-8282. doi: 10.1257/aer.20140260.

Jason Abaluck, Leila Agha, David C Chan Jr, Daniel Singer, and Diana Zhu. Mastering the Art of Cookbook Medicine: Machine Learning, Randomized Trials, and Misallocation. page 71, 2020.

Leila Agha and Dan Zeltzer. Drug Diffusion Through Peer Networks: The Influence of Industry Payments. page w26338, October 2019. doi: 10.3386/w26338.

Ismael Aguilera, Antonio Infante, Héctor Ormeño, and Carlos Urriola. Implementation of the Universal Access with Explicit Guarantees (AUGE) reform Chile case Study. Technical report, World Health Organization, September 2015.

Luis Alcocer, German Novoa, and David Sotres. Digit preferences observed in the measurement of blood pressure: repercussions on the success criteria in current treatment of hypertension. *American Journal of Therapeutics*, 4(9-10):311–314, October 1997. ISSN 1075-2765. doi: 10.1097/00045391-199709000-00005.

Keaven M Anderson, Patricia M Odell, Peter W F Wilson, and B Kannel. Cardiovascular disease risk profiles. 1991.

Katherine Baicker, Amitabh Chandra, Jonathan S. Skinner, and John E. Wennberg. Who You Are And Where You Live: How Race And Geography Affect The Treatment Of Medicare Beneficiaries. *Health Affairs*, 23(Suppl2):VAR–33, January 2004. ISSN 0278-2715. doi: 10.1377/hlthaff.var.33. Num Pages: VAR-44 Publisher: Health Affairs.

Robert B. Baron. Treating Blood Pressure Correctly by Measuring It Correctly. *JAMA Internal Medicine*, 178(6):860–861, June 2018. ISSN 2168-6106. doi: 10.1001/jamainternmed.2018.0311.

Sören Blomquist, Whitney Newey, Anil Kumar, and Che-Yuan Liang. On Bunching and Identification of the Taxable Income Elasticity. Technical Report w24136, National Bureau of Economic

Research, December 2017.

Thomas J. Bossert and Thomas Leisewitz. Innovation and Change in the Chilean Health System. *New England Journal of Medicine*, 374(1):1–5, January 2016. ISSN 0028-4793. doi: 10.1056/NEJMp1514202.

Daniel S. Budnitz, Nadine Shehab, Maribeth C. Lovegrove, Andrew I. Geller, Jennifer N. Lind, and Daniel A. Pollock. US Emergency Department Visits Attributed to Medication Harms, 2017-2019. *JAMA*, 326(13):1299, October 2021. ISSN 0098-7484. doi: 10.1001/jama.2021.13844.

Brantly Callaway, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. Difference-in-Differences with a Continuous Treatment, January 2024. arXiv:2107.02637 [econ].

David C. Chan and Jonathan Gruber. Provider Discretion and Variation in Resource Allocation: The Case of Triage Decisions. *AEA Papers and Proceedings*, 110:279–283, May 2020. ISSN 2574-0768, 2574-0776. doi: 10.1257/pandp.20201032.

David C Chan, Matthew Gentzkow, and Chuan Yu. Selection with Variation in Diagnostic Skill: Evidence from Radiologists*. *The Quarterly Journal of Economics*, page qjab048, January 2022. ISSN 0033-5533. doi: 10.1093/qje/qjab048.

Amitabh Chandra and Douglas O Staiger. EXPERTISE, UNDERUSE, AND OVERUSE IN HEALTHCARE. page 25, 2014.

Raj Chetty, John N. Friedman, Tore Olsen, and Luigi Pistaferri. Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *The Quarterly Journal of Economics*, 126(2):749–804, May 2011. ISSN 0033-5533, 1531-4650. doi: 10.1093/qje/qjr013.

CMS. Medicare Program; Merit-Based Incentive Payment System (MIPS) and Alternative Payment Model (APM) Incentive Under the Physician Fee Schedule, and Criteria for Physician-Focused Payment Models. Final rule with comment period. *Federal Register*, 81(214):77008–77831, November 2016. ISSN 0097-6326.

Stephen Coussens. Heuristic Thinking in the Emergency Department. page 59, 2018.

Janet Currie and Jonathan Zhang. Doing More with Less: Predicting Primary Care Provider Effectiveness. *Review of Economics and Statistics*, 107(2):289–305, March 2025. ISSN 0034-6535, 1530-9142. doi: 10.1162/rest_a_01290. Publisher: MIT Press.

Jishnu Das, Alaka Holla, Aakash Mohpal, and Karthik Muralidharan. Quality and Accountability in Health Care Delivery: Audit-Study Evidence from Primary Care in India. *American Economic Review*, 106(12):3765–3799, December 2016. ISSN 0002-8282. doi: 10.1257/aer.20151138.

Rebecca Diamond and Petra Persson. The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests. Technical Report w22207, National Bureau of Economic Research, May 2016.

Liran Einav, Amy Finkelstein, and Paul Schrimpf. The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D *. *The Quarterly Journal of Economics*, 130 (2):841–899, May 2015. ISSN 0033-5533. doi: 10.1093/qje/qjv005.

Liran Einav, Amy Finkelstein, and Paul Schrimpf. Bunching at the kink: Implications for spending responses to health insurance contracts. *Journal of Public Economics*, 146:27–40, February 2017. ISSN 00472727. doi: 10.1016/j.jpubeco.2016.11.011.

Amy Finkelstein, Matthew Gentzkow, and Heidi Williams. Sources of Geographic Variation in Health Care: Evidence From Patient Migration*. *The Quarterly Journal of Economics*, 131(4): 1681–1726, November 2016. ISSN 0033-5533. doi: 10.1093/qje/qjw023.

Alexander C. Flint, Carol Conell, Xiushui Ren, Nader M. Banki, Sheila L. Chan, Vivek A. Rao, Ronald B. Melles, and Deepak L. Bhatt. Effect of Systolic and Diastolic Blood Pressure on Cardiovascular Outcomes. *New England Journal of Medicine*, 381(3):243–251, July 2019. ISSN 0028-4793. doi: 10.1056/NEJMoa1803180. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa1803180.

Michelle Greiver, Sumeet Kalia, Teja Voruganti, Babak Aliarzadeh, Rahim Moineddin, William Hinton, Martin Dawes, Frank Sullivan, Saddaf Syed, John Williams, and Simon de Lusignan. Trends in end digit preference for blood pressure and associations with cardiovascular outcomes in Canadian and UK primary care: a retrospective observational study. *BMJ Open*, 9 (1):e024970, January 2019. ISSN 2044-6055, 2044-6055. doi: 10.1136/bmjopen-2018-024970. Publisher: British Medical Journal Publishing Group Section: Cardiovascular medicine.

Jonathan Gruber, Thomas P. Hoe, and George Stoye. Saving Lives by Tying Hands: The Unexpected Effects of Constraining Health Care Providers. Technical Report w24445, National Bureau of Economic Research, March 2018.

Helena M. Hernández-Pizarro, Catia Nicodemo, and Guillem López Casasnovas. Discontinuous system of allowances: The response of prosocial health-care professionals. *Journal of Public Economics*, 190:104248, October 2020. ISSN 00472727. doi: 10.1016/j.jpubeco.2020.104248.

Shan Huang and Hannes Ullrich. Physician effects in antibiotic prescribing: Evidence from physician exits. page 62, 2021.

JLN. Case Studies on Payment Innovation for Primary Health Care. Technical report, 2017.

William B. Kannel, Tavia Gordon, and Melvin J. Schwartz. Systolic versus diastolic blood pressure and risk of coronary heart disease: The Framingham study. *The American Journal of Cardiology*, 27(4):335–346, April 1971. ISSN 0002-9149. doi: 10.1016/0002-9149(71)90428-0.

Henrik J. Kleven and Mazhar Waseem. Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan*. *The Quarterly Journal of Economics*, 128(2):669–723, May 2013. ISSN 0033-5533, 1531-4650. doi: 10.1093/qje/qjt004. URL `https://academic.oup.com/qje/article/128/2/669/1943151`.

Henrik Jacobsen Kleven. Bunching. *Annual Review of Economics*, 8(1):435–464, 2016. doi: 10.1146/annurev-economics-080315-015234. _eprint: https://doi.org/10.1146/annurev-economics-080315-015234.

Fernando Lanas, Maria Soledad Garcia, Melanie Paccot, Jimena Peña, Pamela Torres, Thomas Beaney, Anca Chis Ster, Neil R Poulter, and Xin Xia. May Measurement Month 2018: an analysis of blood pressure screening results from Chile. *European Heart Journal Supplements*, 22(Supplement_H):H37–H39, August 2020. ISSN 1520-765X. doi: 10.1093/eurheartj/suaa023.

Panos Mavrokonstantis. Introduction to the bunching Package. page 19, 2019.

Fen Miao, Bo Wen, Zhejing Hu, Giancarlo Fortino, Xi-Ping Wang, Zeng-Ding Liu, Min Tang, and Ye Li. Continuous blood pressure measurement from one-channel electrocardiogram signal using deep-learning techniques. *Artificial Intelligence in Medicine*, 108:101919, August 2020. ISSN 0933-3657. doi: 10.1016/j.artmed.2020.101919.

Katherine T. Mills, Andrei Stefanescu, and Jiang He. The global epidemiology of hypertension. *Nature Reviews Nephrology*, 16(4):223–237, April 2020. ISSN 1759-507X. doi: 10.1038/s41581-019-0244-2. Number: 4 Publisher: Nature Publishing Group.

MINSAL. Guía clínica: hipertensión arterial primaria o esencial en personas de 15 años y más. Technical report, Minsal, Santiago, 2010. OCLC: 1318660964.

MINSAL. Orientación Técnica Programa de Salud Cardiovascular, 2017. URL `https://www.u-cursos.cl/medicina/2017/0/DPPAEF/1/foro/r/OT-PROGRAMA-DE-SALUD-CARDIOVASCULAR_05.pdf`.

Sendhil Mullainathan and Ziad Obermeyer. Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care*. *The Quarterly Journal of Economics*, page qjab046, December 2021. ISSN 0033-5533. doi: 10.1093/qje/qjab046.

Paul Muntner, Shimbo Daichi, Carey Robert M., Charleston Jeanne B., Gaillard Trudy, Misra Sanjay, Myers Martin G., Ogedegbe Gbenga, Schwartz Joseph E., Townsend Raymond R., Urbina Elaine M., Viera Anthony J., White William B., Wright Jackson T., and null null. Measurement of Blood Pressure in Humans: A Scientific Statement From the American Heart Association. *Hypertension*, 73(5):e35–e66, May 2019. doi: 10.1161/HYP.0000000000000087. Publisher: American Heart Association.

Andrew R. Olenski, André Zimerman, Stephen Coussens, and Anupam B. Jena. Behavioral Heuristics in Coronary-Artery Bypass Graft Surgery. *New England Journal of Medicine*, 382(8):778–779, February 2020. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMc1911289.

Yechiam Ostchega and Duong T Nguyen. Hypertension Prevalence Among Adults Aged 18 and

Over: United States, 2017–2018. (364):8, 2020.

PAHO. Chile. 2020. URL https://www.paho.org/salud-en-las-americas-2017/?page_id=105.

Paul Rodríguez-Lesmes and Marcos Vera-Hernández. Multitasking, Two-part Contracts, and Bunching: an Application to Doctors' Jobs and Incentive Contracts. 2021.

Jane Runnacles, Alice Roueché, and Peter Lachman. The right care, every time: improving adherence to evidence-based guidelines. *Archives of Disease in Childhood - Education and Practice*, 103(1):27–33, February 2018. ISSN 1743-0585, 1743-0593. doi: 10.1136/archdischild-2017-312740. Publisher: Royal College of Paediatrics and Child Health Section: E&P: Quality improvement.

Emmanuel Saez. Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy*, 2(3):180–212, August 2010. ISSN 1945-7731. doi: 10.1257/pol.2.3.180.

Manasvini Singh. Heuristics in the delivery room. *Science*, 374(6565):324–329, October 2021. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abc9818.

Matthew Stephens, Peter Carbonetto, David Gerard, Mengyin Lu, Lei Sun, Jason Willwerscheid, and Nan Xiao. ashr: Methods for Adaptive Shrinkage, using Empirical Bayes, December 2016. URL https://CRAN.R-project.org/package=ashr. Institution: Comprehensive R Archive Network Pages: 2.2-63.

George S. Stergiou, Eoin O'Brien, Martin Myers, Paolo Palatini, Gianfranco Parati, and the STRIDE BP Scientific Advisory Board. STRIDE BP: an international initiative for accurate blood pressure measurement. *Journal of Hypertension*, 38(3):395–399, March 2020. ISSN 0263-6352. doi: 10.1097/HJH.0000000000002289.

Stefan Timmermans and Aaron Mauck. The Promises And Pitfalls Of Evidence-Based Medicine. *Health Affairs*, 24(1):18–28, January 2005. ISSN 0278-2715. doi: 10.1377/hlthaff.24.1.18. Publisher: Health Affairs.

Amos Tversky and Daniel Kahneman. Judgment under Uncertainty: Heuristics and Biases. 185:

10, 1974.

Bernard Vrijens, Sotiris Antoniou, Michel Burnier, Alejandro de la Sierra, and Massimo Volpe. Current Situation of Medication Adherence in Hypertension. *Frontiers in Pharmacology*, 8, March 2017. ISSN 1663-9812. doi: 10.3389/fphar.2017.00100.

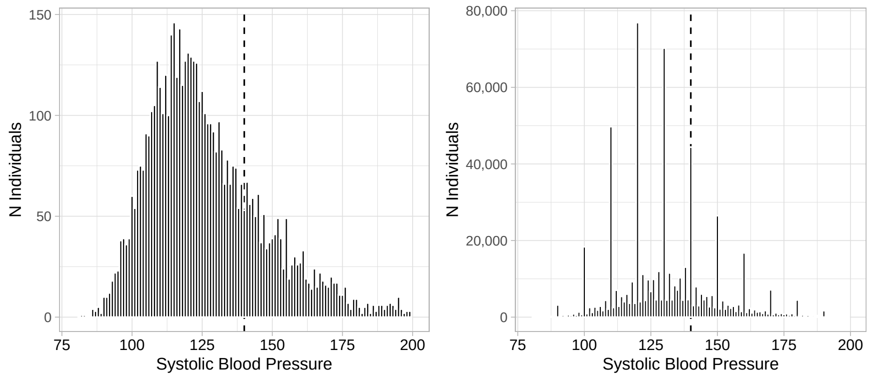
WHO. A global brief on Hypertension: Silent killer, global public health crisis. Technical report, 2013. URL `https://apps.who.int/iris/bitstream/handle/10665/79059/W?sequence=1`.

Jane Wilcox. Observer Factors in the Measurement of Blood Pressure. *Nursing Research*, 10(1): 4–17, January 1961. ISSN 0029-6562.

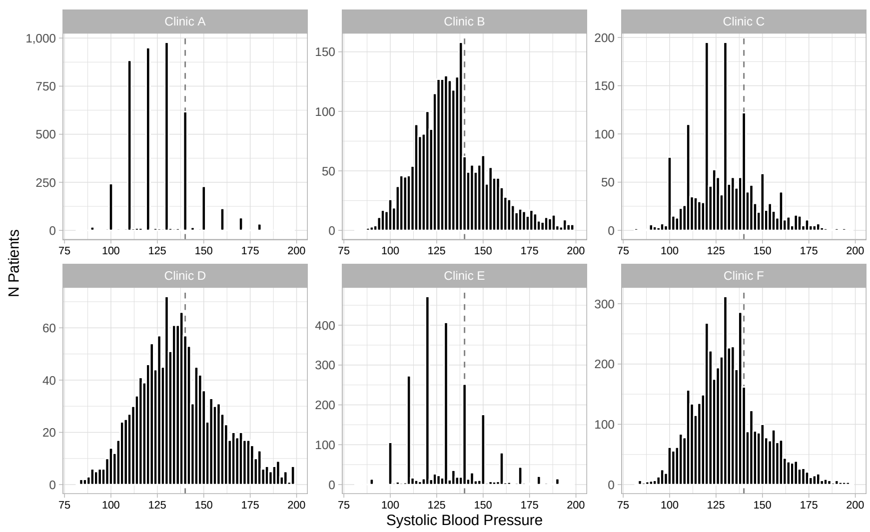
8 Tables and Figures

Figure 1: Expected vs. Observed Distribution of Systolic Blood Pressure

Panel A: National Health Survey **Panel B**: Primary Care Records, Overall



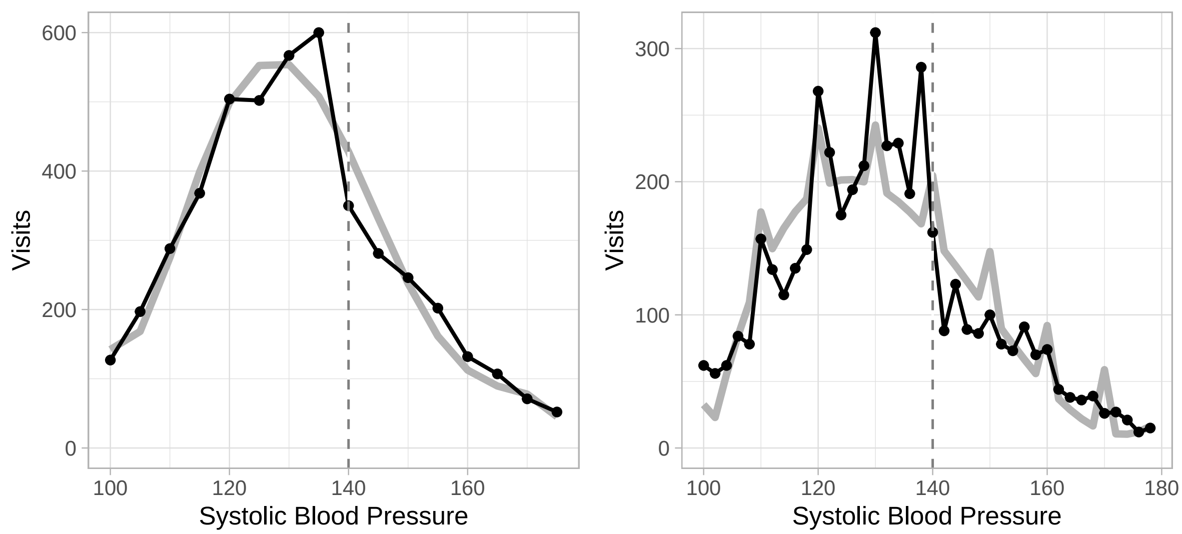
Panel C: Primary Care Records, Six Clinics



Note: Comparison of expected (panel A) to observed (panel B) distribution of systolic blood pressure in two large populations. Panel A shows the distribution of systolic blood pressure from Chile's nationally representative health survey (ENS 2016-2017). Panel B shows the distribution of systolic blood pressure from primary care electronic health records where substantial end-digit bias, or rounding to zero, is seen. Panel C shows the distributions of systolic blood pressure from 6/257 primary care clinics from preventative care visits among patients who were not previously diagnosed with hypertension. Clinics A, C, and E show substantial left-digit bias: numbers that end in zero are over-represented. Clinics A, B, and F show substantial bunching: there is missing mass above 140 and excess mass below it.

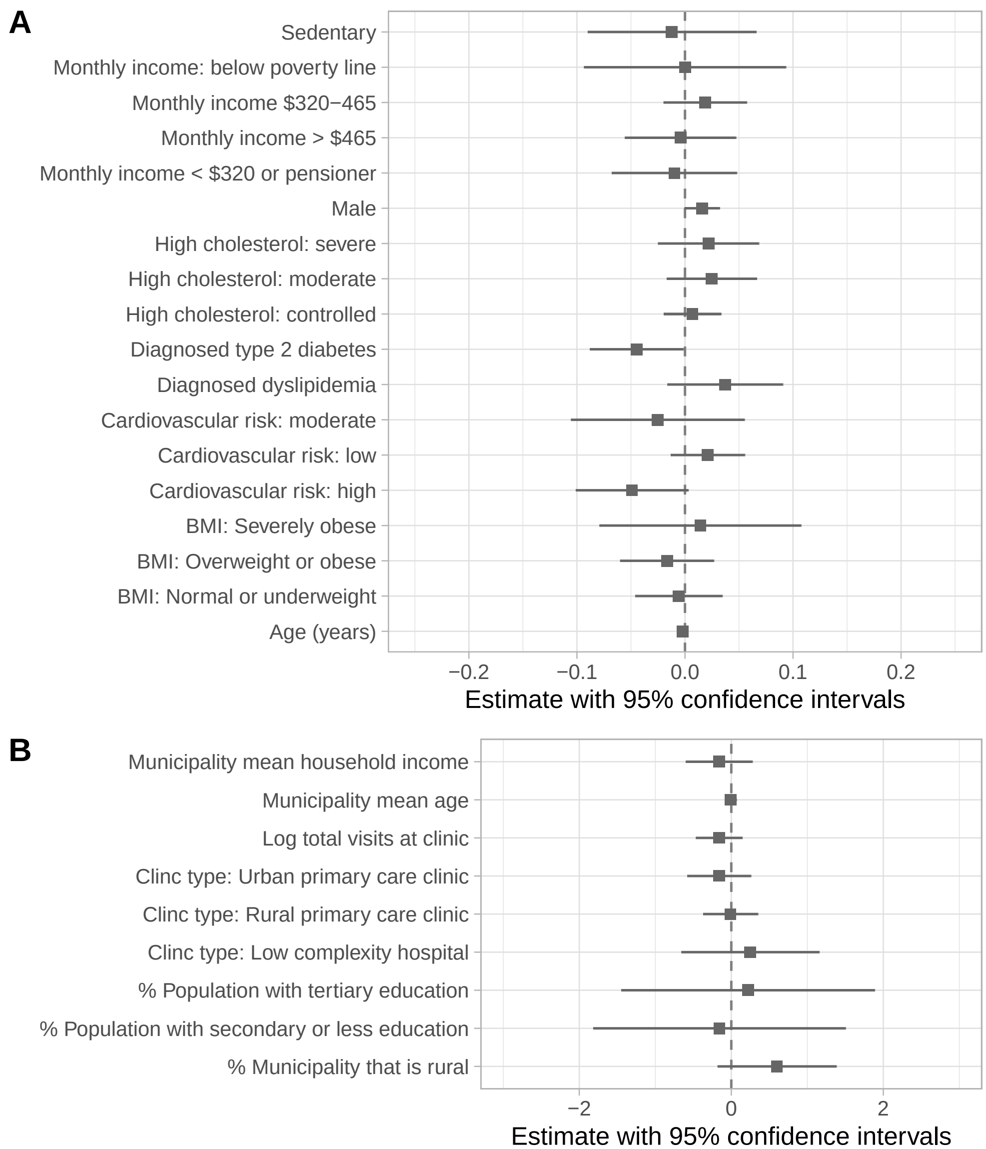
Figure 2: Example of Counterfactual Bunching Estimation

Panel A: Clinic with Little Rounding **Panel B**: Clinic with Substantial Rounding



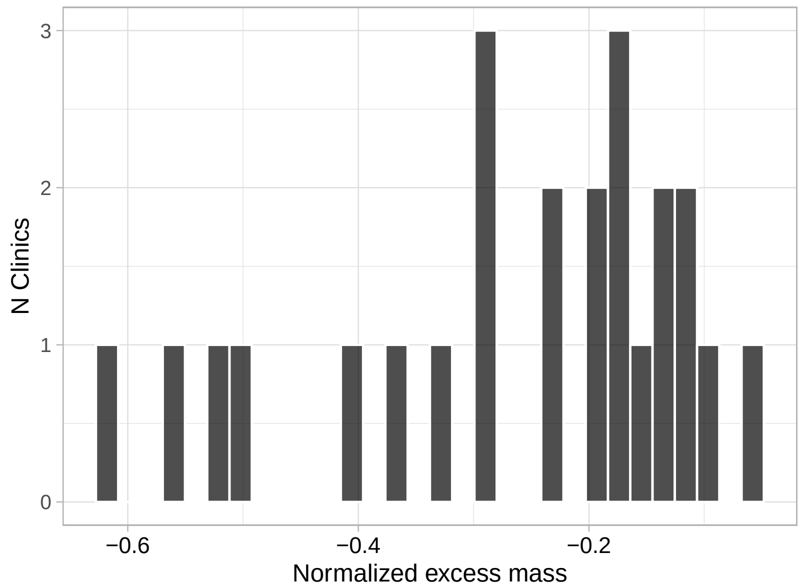
Note: Examples of bunching estimation. The grey line is the counterfactual density from equation (1). Panel A shows a clinic with little rounding to zero. Panel B shows a clinic with more rounding to zero, seen as spikes at 110, 120, 130,..., 150. The dashed line at 140 shows the diagnostic threshold. Estimated using the bunching package in R [Mavrokonstantis, 2019].

Figure 3: Balance of Patient and Clinic Characteristics by Magnitude of Discretion



Note: This figure tests for correlation between patient, clinic, and municipality characteristics and the magnitude of discretion. Point estimates and 95% confidence intervals shown from the regression of clinics' continuous and standardized bunching magnitude on (A) patient characteristics, and (B) clinic and municipality characteristics. Each characteristic is from a separate regression and includes year and quarter fixed effects, with standard errors clustered at the clinic level.

Figure 4: Estimated Magnitude of Discretion in Blood Pressure Measurement



Note: Histogram of the magnitude of discretion, estimated using bunching, for clinics with statistically significant negative bunching (24/257). Negative values indicate missing mass above the threshold, and excess mass below it. Each clinic's excess mass estimate was normalized by dividing the total excess mass by the height of the counterfactual at the threshold.

Table 1: Summary Statistics

Patient Characteristics	Mean	St. Dev.
Female	0.618	0.486
Mean age	57.711	13.087
Age 18-29	0.031	0.173
Age 30-39	0.069	0.253
Age 40-49	0.172	0.377
Age 50-59	0.263	0.440
Age 60-69	0.264	0.441
Age 70-79	0.201	0.401
Body mass index	30.714	6.013
Normal or underweight BMI	0.211	0.408
Overweight BMI	0.375	0.484
Obese BMI	0.482	0.500
Severely obese BMI	0.067	0.251
Type 2 Diabetes	0.286	0.452
Dyslipidemia	0.641	0.480
High cholesterol	0.351	0.477
Cardiovascular risk: high	0.271	0.445
Cardiovascular risk: moderate	0.168	0.374
Cardiovascular risk: low	0.129	0.335
Sedentary	0.557	0.497
Monthly income: below poverty line	0.267	0.442
Monthly income <$320 or pensioner	0.435	0.496
Monthly income $320-465	0.122	0.327
Monthly income >$465	0.171	0.376

Municipality & Clinic Characteristics	Mean	St. Dev.
Household income per capita	440.827	229.015
Log visits per clinic	7.938	0.460
Mean municipality age	36.521	2.343
Secondary or primary education	0.865	0.088
% of municipality that is rural	0.148	0.173
Tertiary education	0.135	0.088
Urban primary care clinic	0.825	0.381
Rural primary care clinic	0.089	0.286
Low complexity hospital	0.058	0.235
Literacy rate in municipality	0.966	0.024
Employment rate in municipality	0.525	0.056

Note: Panel A: Mean and standard deviation (St. Dev.) of patient-level characteristics for 619,907 patients. BMI stands for body mass index. Cardiovascular risk is assigned by the healthcare provider using a modified Framingham 10-year cardiovascular risk algorithm. Monthly income categories come from the patient's FONASA insurance level. All variables are indicators except for age and body mass index. Panel B: Mean and standard deviation of clinic- and municipality-level characteristics for 257 clinics. Municipality-level variables come from Chile's Socio-economic Survey, and are linked to clinics based on their location. Household income per capita is monthly a municipality mean, and in 2020 USD.

Table 2: Impact of Discretion on Hypertension Diagnosis and Prescription

	Hypertension Diagnosis	Hypertension Prescription
	(1)	(2)
Discretion x BP<140	−0.006	0.008
	(0.006)	(0.010)
Discretion	−0.009	−0.021
	(0.006)	(0.014)
BP<140	−0.251***	−0.169***
	(0.005)	(0.005)
Observations	619,907	619,907
Mean dep. var.	0.522	0.376
Clinics	257	257

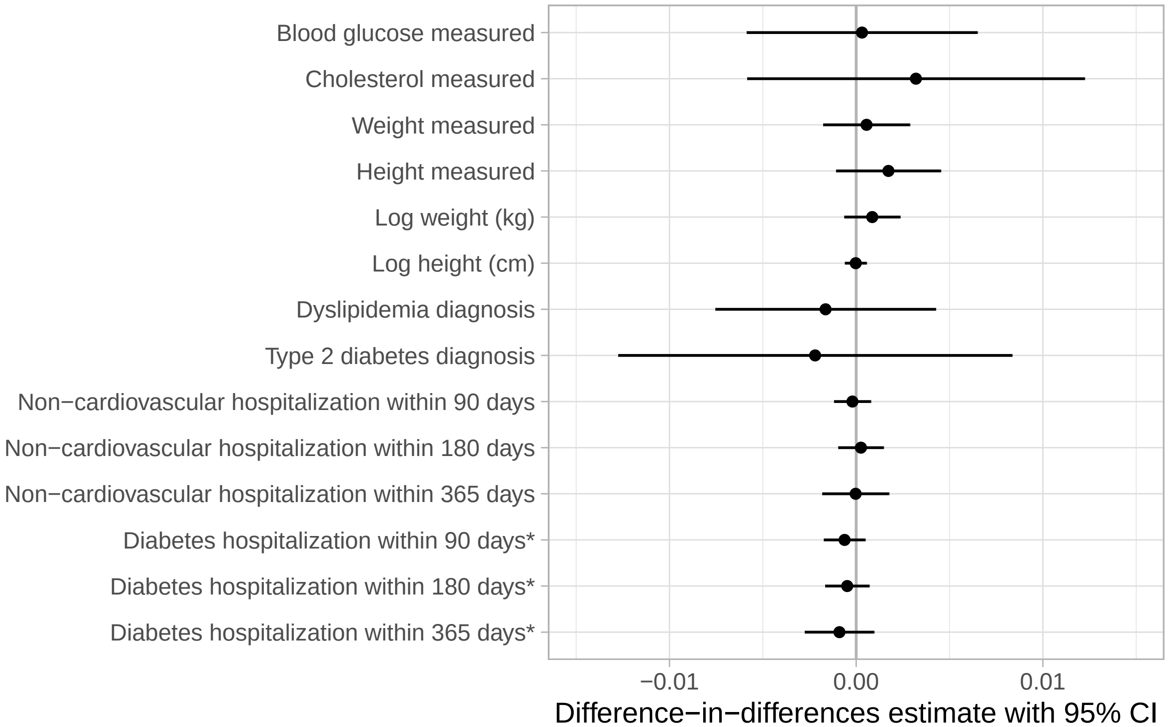
Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences results estimated using equation (2). BP< 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of bunching at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for year and quarter of primary care visit. Hypertension diagnosis is an indicator for if the patient was newly diagnosed. Hypertension prescription is an indicator for if the patient was prescribed a new hypertension medication.

Table 3: Impact of Discretion on Cardiovascular Hospitalization 3, 6, and 12 Months After Primary Care Visit

	Stroke			Heart Attack			Congestive Heart Failure		
	≤3 mo.	≤6 mo.	≤12 mo.	≤3 mo.	≤6 mo.	≤12 mo.	≤3 mo.	≤6 mo.	≤12 mo.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Discretion x BP<140	-0.018^{**}	-0.039^{***}	-0.061^{***}	-0.028^{**}	-0.040^{***}	-0.035^{*}	0.003	0.002	-0.0003
	(0.009)	(0.012)	(0.019)	(0.012)	(0.013)	(0.019)	(0.006)	(0.010)	(0.013)
Discretion	0.018^{**}	0.039^{***}	0.059^{***}	0.022^{*}	0.033^{**}	0.029	-0.005	-0.009	-0.012
	(0.009)	(0.012)	(0.017)	(0.012)	(0.014)	(0.020)	(0.005)	(0.008)	(0.010)
BP<140	-0.029^{***}	-0.061^{***}	-0.107^{***}	-0.025^{***}	-0.062^{***}	-0.119^{***}	0.010	0.013	0.004
	(0.005)	(0.009)	(0.014)	(0.009)	(0.013)	(0.019)	(0.007)	(0.010)	(0.015)
Observations	619,907	619,907	619,907	619,907	619,907	619,907	619,907	619,907	619,907
Mean dep. var.(%)	0.029	0.066	0.137	0.109	0.205	0.385	0.062	0.124	0.232
Clinics	257	257	257	257	257	257	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences results estimated using equation (2). BP< 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of bunching at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. All dependent variables in this table are multiplied by 100. Each dependent variable is an indicator for if the patient was hospitalized with the listed condition, within 3, 6 or 12 months of their primary care visit, zero otherwise. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for male, 1-year age, year and quarter of primary care visit. Stroke includes any cerebral infarction (ICD-10 codes I63). Heart attack includes acute coronary syndrome, myocardial infarction, and any other condition associated with sudden, reduced blood flow to the heart (I20-22, I24-25). Congestive heart failure is I50.

Figure 5: Impact of Discretion on Placebo Outcomes



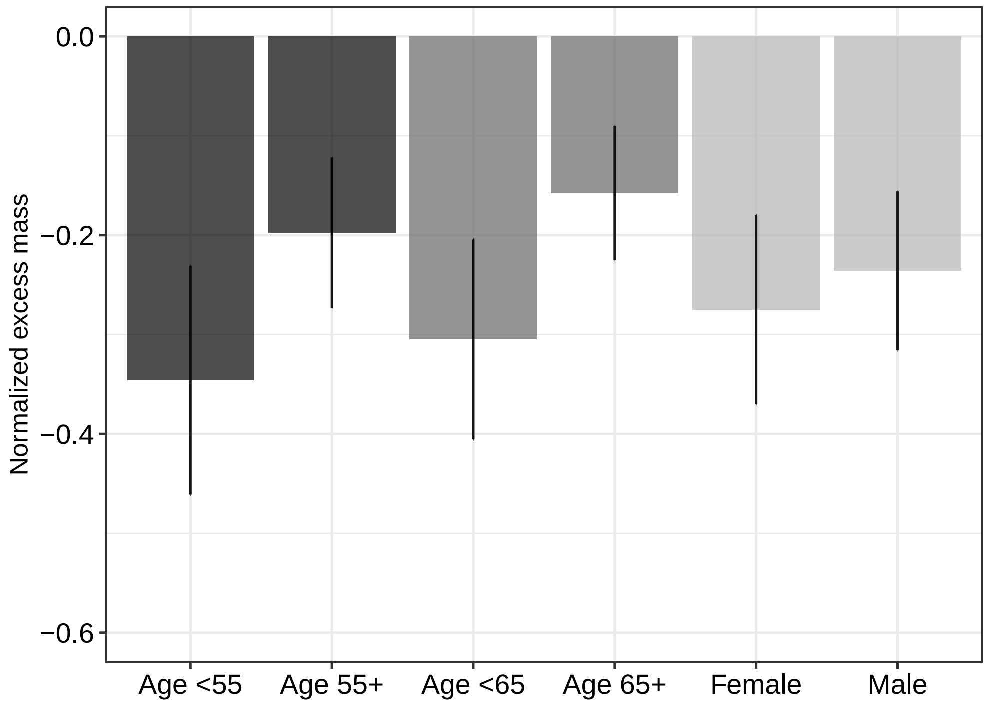
Note: This figure presents difference-in-differences results estimated using equation (2). β_3, the coefficient on the interaction term between discretion and BP$<$140 is plotted, along with 95% confidence intervals. Outcomes other than height and weight are indicators. The asterisk indicates the outcome is defined only among patients with a prior type 2 diabetes diagnosis (177,083 observations). Corresponding tables: A8, A9.

Table 4: Selection by Patient Characteristics Representative of High Cardiovascular Risk

	Male	Severe Obesity	Age 55+	Age 65+
	(1)	(2)	(3)	(4)
Discretion x BP<140	-0.006^{***}	0.001	-0.006^{**}	-0.005^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
Discretion	0.008^{***}	-0.0002	-0.006	-0.008
	(0.002)	(0.004)	(0.005)	(0.005)
BP<140	-0.096^{***}	-0.019^{***}	-0.141^{***}	-0.127^{***}
	(0.002)	(0.001)	(0.004)	(0.003)
Observations	619,907	582,880	619,907	619,907
Mean dep. var.	0.382	0.067	0.596	0.336
Clinics	257	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences results estimated using equation (2). BP< 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of bunching at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Each dependent variable is an indicator for if the patient has the characteristic listed, zero otherwise. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for year and quarter of primary care visit.

Figure 6: Estimated Magnitude of Discretion in Blood Pressure Measurement by Patient Characteristics



Note: This figure presents the magnitude of discretion (bunching) in blood pressure among high discretion clinics, defined as clinics with an overall magnitude of discretion > 0 at $p < 0.05$. Normalized excess mass was obtained by conditioning on patient characteristic (age, sex), and conducting bunching estimation on the distribution of systolic blood pressure using equation (1). Bunching is more pronounced among patients with higher cardiovascular risk factors (male, age $\geq$55, age $\geq$65) than among those with lower cardiovascular risk (female, age <55, age <65). However, within-group differences are statistically significant only for age 65 (p<0.05).

Online Appendix

For Discretion in Clinical Decision Making: Evidence from Bunching by Claire Boone

Contents

A.1 Sample Selection

Table A1: Inclusion Criteria for Primary Care Electronic Health Records Data

Inclusion criteria	**N patients**	**N clinics**	**% patients excluded**
-	1,448,800	552	-
Blood pressure was measured	1,413,499	545	2.4%
No previous hypertension diagnosis or medication prescription	955,528	539	32.4%
Age 18-79	863,571	538	9.6%
Large clinics	755,653	257	12.5%
At least one year of follow-up time	619,907	257	18.0%

Note: Inclusion criteria and number and percent of patients excluded at each step. Large is defined a clinic above the median number of patient-visits (N=1265). The final sample is 619,907 patient-visits and 257 clinics.

Table A2: Patient Reasons for Admission to Cardiovascular Health Program

	N	%
History of CVD	393,657	63.5
Dyslipidemia	285,304	46.0
Type 2 diabetes	177,083	28.6
Smoker and over 55	9,643	1.6
Unknown	52	0.0
Unique patients	619,907	

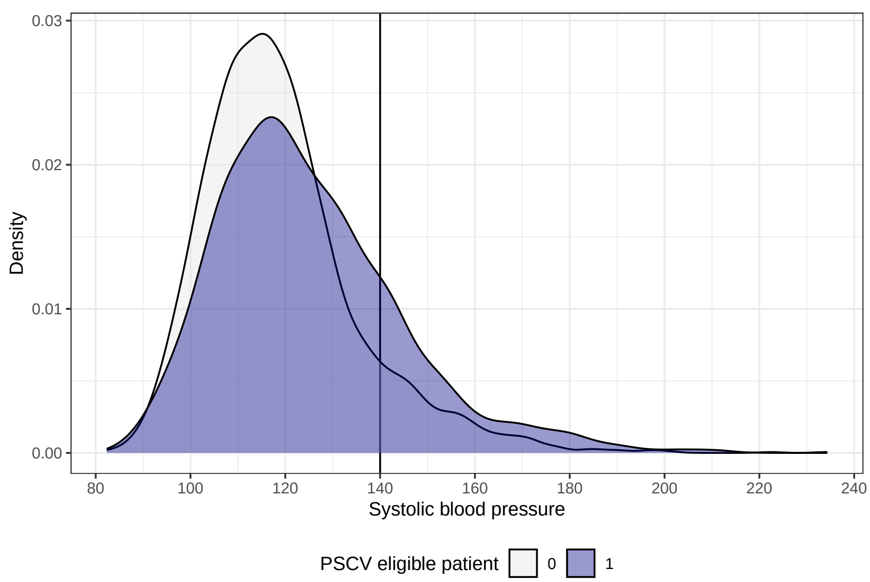
Note: Patients are enrolled in Chile's Cardiovascular Health Program (*Programa Salud Cardiovascular*) if they have one or more of the characteristics in the table, or hypertension. Categories are not mutually exclusive. Patients in the Cardiovascular Program with previously diagnosed hypertension were excluded.

A.1.1 PSCV vs. non PSCV patients

Only EHR data from PSCV eligible patients is available. Could missing mass in the distribution of blood pressure be caused by non-PSCV patients? First, PSCV patients are defined as having high cardiovascular risk, and therefore are likely to have *higher* blood pressure than non-PSCV patients. Excluding non-PSCV patients is therefore unlikely to cause missing mass in the regions I observe: approximately 140-150 mmHg. While I cannot check whether this is true using my primary dataset, I can check it using Chile's National Health Survey (ENS).

To do this, I classified ENS participants into PSCV eligible or PSCV non-eligible. Patients are eligible for PSCV if they have type 2 diabetes, dyslipidemia, or have risk factors: a history of cardiovascular disease (CVD), or smoke and are at least 55 years old [MINSAL, 2017]. While prior diagnosis with hypertension is an inclusion criteria for PSCV, these patients were excluded from both my paper's analysis and from the ENS sample. The ENS does not ask about every disease that is considered in a patient's history of cardiovascular disease by the PSCV program[4] but I believe this is still a useful comparison. Figure A1 shows that among PSCV eligible patients, the distribution of systolic blood pressure is shifted right compared to non-eligible patients, suggesting that missing mass above 140 mmHg is unlikely to be caused by using only PSCV patients in my analysis.

Figure A1: Blood pressure among PSCV-eligible vs. non-eligible patients in the National Health Survey



Note: Distribution of systolic blood pressure from Chile's National Health Survey (ENS), comparing patients eligible for the PSCV program vs. not.

[4]History of CVD defined by the PSCV program includes patient self report of: Coronary heart disease: Acute myocardial infarction, stable/unstable angina, history of angioplasty or myocardial revascularization surgery; Cerebrovascular disease: Stroke or transient ischemic stroke; Peripheral arterial disease; Atherosclerotic aortic disease; Renovascular disease; Carotid disease. The National Health Survey only asked participants about their history of Coronary heart disease and Cerebrovascular disease.

Table A3: Comparing Mean Patient Characteristics Between Included and Excluded Clinics

Variable	Excluded Clinics	Included Clinics	Mean Diff.	p-val
Male	0.39	0.39	-0.01	0.08
Age	57.18	57.31	0.14	0.66
Type 2 Diabetes	0.21	0.28	0.08	0.00
BMI: Normal	0.12	0.14	0.02	0.00
BMI: Overweight	0.35	0.37	0.02	0.00
BMI: Obese	0.52	0.49	-0.03	0.00
Waist: Obese	0.70	0.69	-0.00	0.76
Cholesterol: normal	0.66	0.65	-0.01	0.08
Cholesterol: high level 1	0.21	0.21	0.00	0.74
Cholesterol: high level 2	0.09	0.10	0.01	0.06
Cholesterol: high level 3	0.04	0.04	0.00	0.04
Sedentary	0.53	0.55	0.02	0.07
Fonasa A	0.33	0.27	-0.05	0.00
Fonasa B	0.41	0.43	0.02	0.04
Fonasa C	0.11	0.12	0.01	0.00
Fonasa D	0.14	0.17	0.02	0.00
Municipality mean log income	5.90	6.02	0.12	0.00
Municipality mean age	36.97	36.52	-0.45	0.06
Municipality share secondary edu. or less	0.90	0.86	-0.03	0.00
Municipality share tertiary edu. or more	0.10	0.14	0.03	0.00
Municipality share rural	0.28	0.15	-0.13	0.00
Log visits at clinic	5.85	7.94	2.09	0.00
Urban clinic	0.47	0.82	0.35	0.00
Rural clinic	0.14	0.09	-0.05	0.06
Low complexity hospital	0.09	0.06	-0.03	0.17
N Clinics	239.00	257.00		

Note: This table compares mean patient characteristics between patients age 18-79 and followed for at least one year at large clinics (final analysis sample of 619,907 patients at 257 clinics), to patients age 18-79 and followed for at least one year at small clinics. Large clinics had at least 1265 visits (median). Note that this table averages within clinics, whereas table 1 is at the patient level.

A.2 Additional Institutional Details

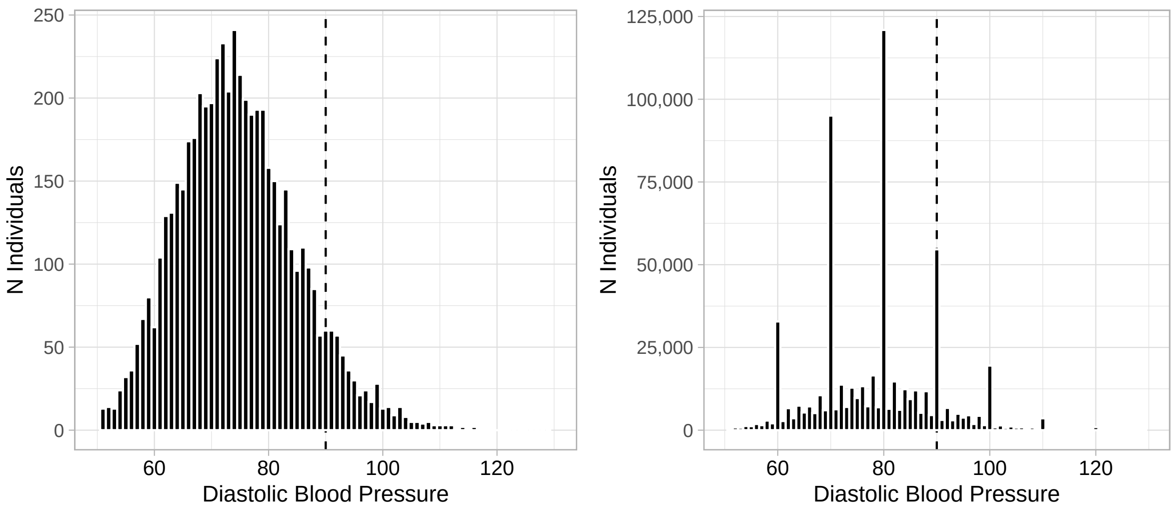
Table A4: Chile's Cardiovascular Management Handbook: Frequent errors that result in an inaccurate measurement of blood pressure

	Discrepancy in Blood Pressure	
Factor	**Systolic**	**Diastolic**
Did not rest 5 minutes prior to measurement	+10-20	0
Talking or active listening during measurement	+ 10-17	+ 10-13
Blood pressure cuff over clothing	+5-40	0
Blood pressure cuff too tight	+10-15	+2-8
Blood pressure cuff too loose	-7	5
Ate, smoked, or drank coffee 30 minutes before measurement	+6-20	0
Back was not against the chair	+6-10	0
Crossed legs	+5-8	0

Note: This table is the author's own translation of appendix table 3 from Chile's 2017 handbook on cardiovascular disease management (Orientación Técnica Programa de Salud Cardiovascular 2017, page 37, MINSAL [2017]). Discrepancy indicates the change in blood pressure relative to the patient's theoretical 'true' blood pressure.

A.3 Diastolic Blood Pressure

Figure A2: Expected vs. Observed Distribution of Diastolic Blood Pressure
Panel A: Chile's Nationally Representative Health Survey. **Panel B**: Primary Care Records



Note: Comparison of expected (panel A) to observed (panel B) distribution of diastolic blood pressure in two large populations. Panel A shows the distribution of diastolic blood pressure from Chile's nationally representative health survey (ENS 2016-2017). Panel B shows the distribution of diastolic blood pressure from primary care electronic health records. Individuals with a previous diagnosis of hypertension were excluded.

A.4 Construction of Variables

Table A5: Construction of Hospitalization Variables

Variable	Description	ICD-10 codes
Stroke	Any cerebral infarction	I63
Heart attack	Also called acute coronary syndrome. Includes myocardial infarction, and other range of conditions associated with sudden, reduced blood flow to the heart	I20, I21, I22, I24, I25
Congestive heart failure	Failure of the heart of pump or fill adequately	I50
Non-cardiovascular	Hospitalization for any reason except stroke, heart attack, and heart failure	Any except I20, I21, I22, I24, I25, I50, I63
Type 2 Diabetes	Hospitalization for reasons directly related to type 2 diabetes	E10, E11, I12, N03, N04, N08, N18, Z79, Z99

Note: This table describes how hospitalization variables were constructed. Each variable is an indicator equal to one if the patient was hospitalized with a primary or secondary diagnosis listed in the ICD-10 column.

A.5 Robustness: Bunching Estimation

Table A6: Robustness of bunching estimation to different specifications: clinics with low share of excess zeros (little end digit bias)

Specifications	**Preferred**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
N clinics	43	21	81	43	43	43	43	43
Low rounding classified as clinics with excess zeros	<10%	<5%	<20%	<10%	<10%	<10%	<10%	<10%
Exclusion window (left of threshold)	0 bins	0 bins	0 bins	0 bins	0 bins	0 bins	0 bins	0 bins
Exclusion window (right of threshold)	0 bins	0 bins	0 bins	1 bin	0 bins	0 bins	0 bins	0 bins
Bin size	5	5	5	5	4	2	5	5
N bins left (100-139; sums to 40)	8	8	8	8	10	20	8	8
N bins right (140-180; sums to 40)	8	8	8	8	10	20	8	8
Polynomial	9	9	9	9	9	9	7	10
Results								
N clinics with negative bunching	43	17	59	35	28	25	30	31
N clinics with statistically significant negative bunching (p<0.05)	18	10	25	17	8	5	12	18
Magnitude of bunching among clinics with significant bunching (p<0.05)								
Mean	-0.19	-0.20	-0.19	-0.43	-0.29	-0.38	-0.23	-0.21
Median	-0.17	-0.19	-0.17	-0.45	-0.26	-0.32	-0.22	-0.18
Minimum	-0.38	-0.38	-0.38	-0.72	-0.53	-0.73	-0.39	-0.39
Maximum	-0.06	-0.06	-0.06	-0.21	-0.12	-0.23	-0.12	-0.10

Note: Table reports estimates of the magnitude of bunching under alternative specifications (columns). First, clinics are classified into “high rounders" or “low rounders" depending on the share of excess zeros observed in their distributions (5-20%), and this table includes low rounding clinics only. Polynomial refers to the numeric value for the order of polynomial for counterfactual fit. All specifications are estimated on systolic blood pressures 100-180 with 40 units on either side of the threshold (140). All specifications include fixed effects at numbers than end in zero, other than 140 (80, 90, ..., 180) to account for rounding.

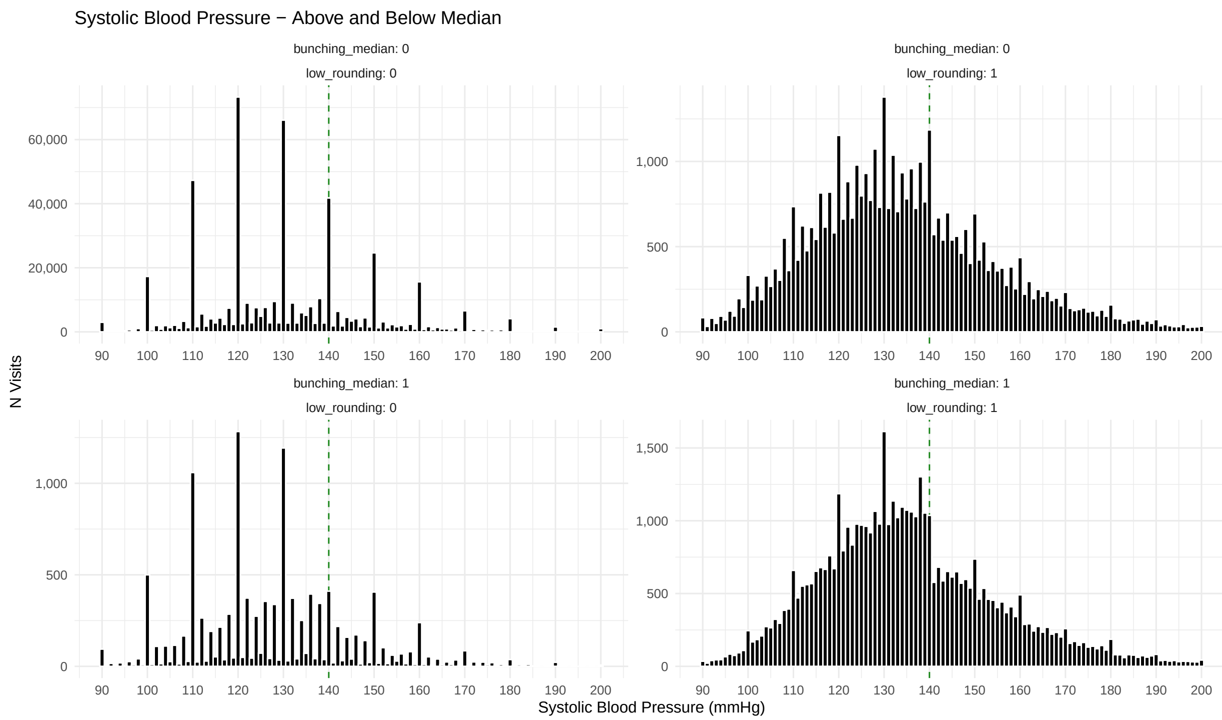
Table A7: Robustness of bunching estimation to different specifications: clinics with high share of excess zeros (substantial end digit bias)

Specifications	**Preferred**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**
N clinics	221	243	183	221	221	221	221	221
Low rounding classified as clinics with excess zeros	≥10%	≥5%	≥20%	≥10%	≥10%	≥10%	≥10%	≥10%
Exclusion window (left of threshold)	0 bins	0 bins	0 bins	0 bins	0 bins	0 bins	0 bins	0 bins
Exclusion window (right of threshold)	0 bins	0 bins	0 bins	2 bins	0 bins	0 bins	0 bins	0 bins
Bin size	2	2	2	2	1	5	2	2
N bins left (100-139; sums to 40)	20	20	20	20	40	8	20	20
N bins right (140-180; sums to 40)	20	20	20	20	40	8	20	20
Polynomial	9	9	9	9	9	9	7	10
Results								
N clinics with negative bunching	109	123	93	66	86	186	118	133
N clinics with statistically significant negative bunching (p<0.05)	6	8	5	3	4	8	5	6
Magnitude of bunching among clinics with significant bunching (p<0.05)								
Mean	-0.49	-0.45	-0.52	-0.66	-0.52	-0.2	-0.51	-0.45
Median	-0.51	-0.51	-0.52	-0.53	-0.53	-0.19	-0.53	-0.47
Minimum	-0.62	-0.62	-0.62	-0.96	-0.65	-0.33	-0.62	-0.62
Maximum	-0.3	-0.22	-0.4	-0.48	-0.37	-0.15	-0.31	-0.26

Note: Table reports estimates of the magnitude of bunching under alternative specifications (columns). First, clinics are classified into “high rounders" or “low rounders" depending on the share of excess zeros observed in their distributions (5-20%), and this table includes high rounding clinics only. Polynomial refers to the numeric value for the order of polynomial for counterfactual fit. All specifications are estimated on systolic blood pressures 100-180 with 40 units on either side of the threshold (140). All specifications include fixed effects at numbers than end in zero, other than 140 (80, 90, ..., 180) to account for rounding.

A.6 Bunching in Blood Pressure and Distributions of Placebo Biomarkers

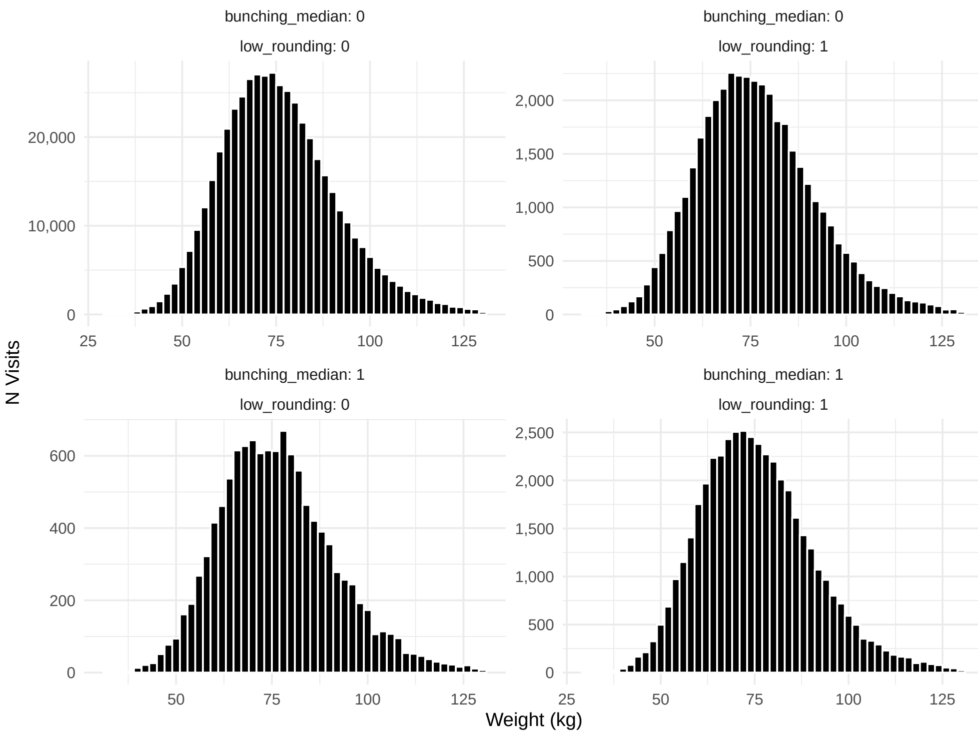
Figure A3: Histogram of Systolic Blood Pressure by High vs. Low Discretion and High vs. Low Rounding Clinics



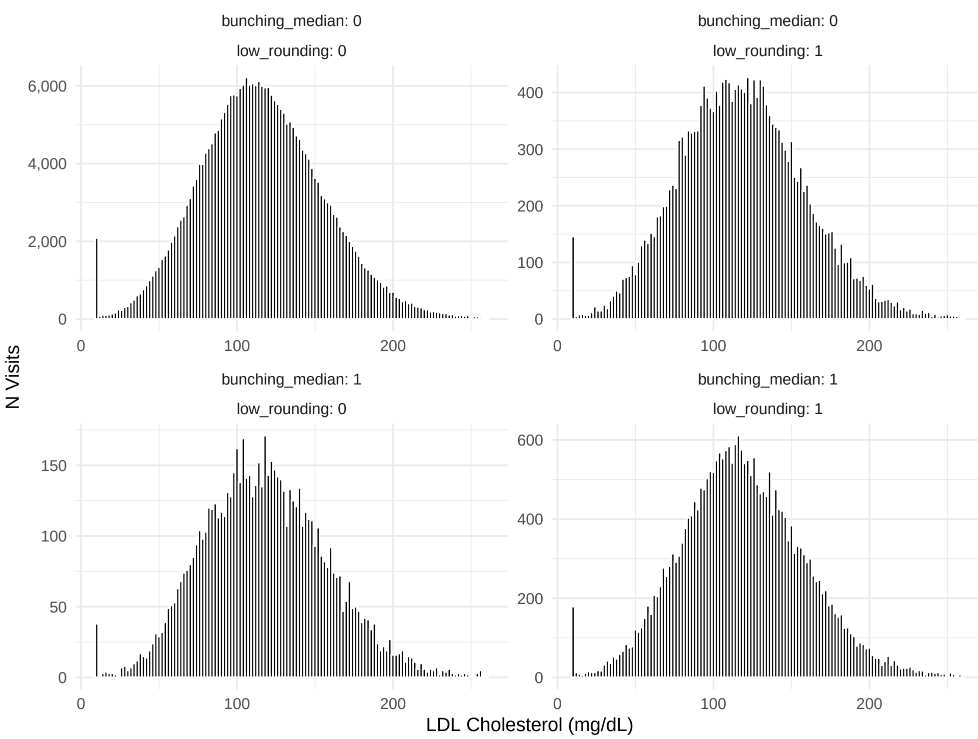
Note: Histogram of systolic blood pressure in EHR data by whether the estimated magnitude of discretion is above or below the median among high or low rounding clinics.

Figure A4: Bunching Not Observed in Placebo Biomarkers by High vs. Low Discretion and High vs. Low Rounding Clinics

Panel A: Weight



Panel B: LDL Cholesterol

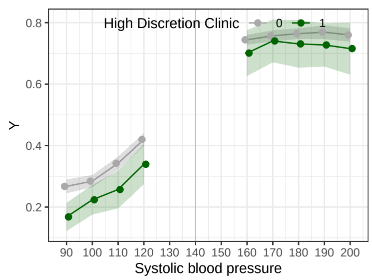


Note: Histogram of weight (panel A) and LDL cholesterol (panel B) in EHR data by whether the estimated magnitude of discretion is above or below the median among high or low rounding clinics.

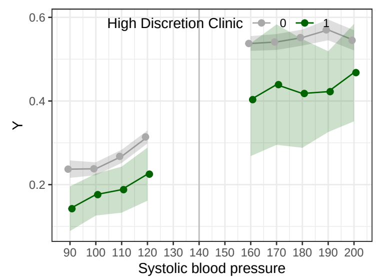
A.7 Mean outcomes by blood pressure bin

Figure A5: Trends in Outcomes by Blood Pressure Bins

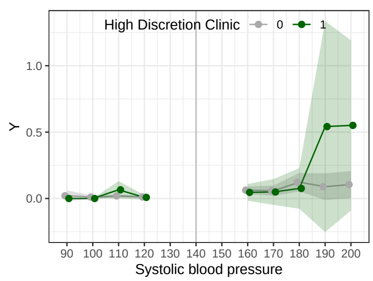
(a) Htn. Diagnosis



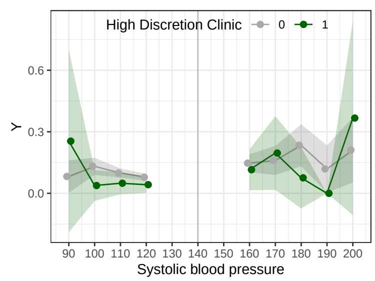
(b) Htn. Prescription



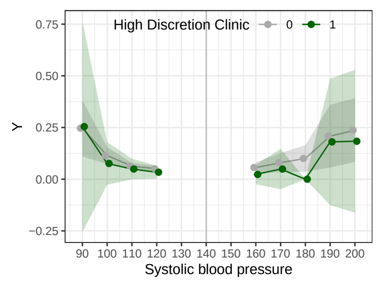
(c) Stroke≤3 months



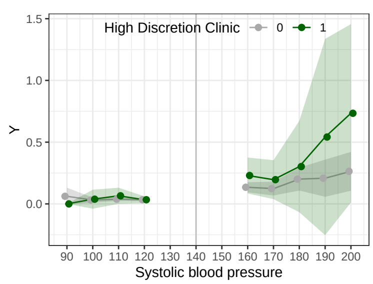
(d) Heart attack≤3 months



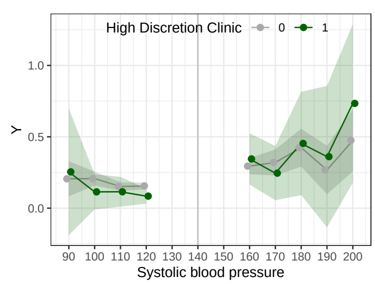
(e) Heart failure≤3 months



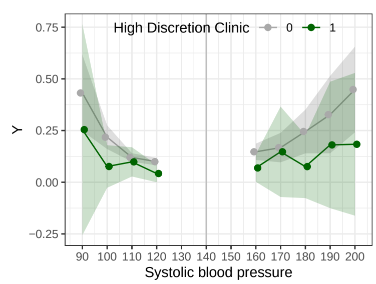
(f) Stroke≤6 months



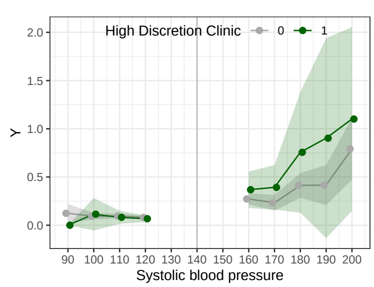
(g) Heart attack≤6 months



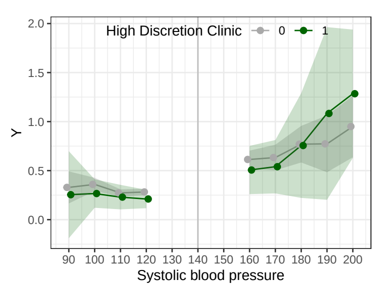
(h) Heart failure≤6 months



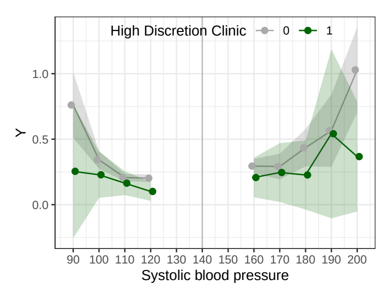
(i) Stroke≤12 months



(j) Heart attack≤12 months



(k) Heart failure≤12 months



Note: Unadjusted mean outcomes by 10-unit systolic blood pressure bin, at clinics with high vs. low discretion. The range of blood pressure where discretion is likely applied, bins 130-150, are excluded as mean outcomes are expected to differ over this range.

A.8 Placebo Outcomes

Table A8: Impact of Discretion on Placebo Outcomes: Hospitalizations 3, 6, and 12 Months After Primary Care Visit

	Non-cardiovascular hospitalization			DM2 hospitalization, among diagnosed		
	$\leq$3 mo.	$\leq$6 mo.	$\leq$12 mo.	$\leq$3 mo.	$\leq$6 mo.	$\leq$12 mo.
	(1)	(2)	(3)	(4)	(5)	(6)
Discretion x BP<140	−0.0002	0.0003	−0.00003	−0.0004	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Discretion	−0.00001	−0.0004	0.0002	0.001	0.0001	−0.002
	(0.0005)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
BP<140	0.0001	−0.001*	−0.003***	0.0001	−0.002*	−0.005***
	(0.0005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Observations	619,907	619,907	619,907	177,083	177,083	177,083
Mean dep. var.(%)	0.024	0.046	0.086	0.027	0.051	0.096
Clinics	257	257	257	254	254	254

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences results estimated using equation (2) on placebo outcomes. BP< 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of bunching at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. DM2 stands for type 2 diabetes. Each dependent variable is an indicator for if the patient was hospitalized with the listed condition, within 3, 6 or 12 months of their primary care visit, zero otherwise. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for male, 1-year age, year and quarter of primary care visit.

Table A9: Impact of Discretion on Placebo Outcomes: Clinical Decisions and Biomarkers

	Tests done at encounter:				Biomarkers:		New diagnosis of:	
	Blood glucose	Cholesterol	Weight	Height	Log weight	Log height	Dyslipidemia	Type 2 diabetes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Discretion x BP<140	0.0003	0.003	0.001	0.002	0.001	−0.0003	−0.001	−0.003
	(0.003)	(0.005)	(0.001)	(0.001)	(0.001)	(0.0004)	(0.003)	(0.005)
Discretion	0.015	0.005	0.003	0.003	−0.0004	0.001	0.009	−0.006
	(0.009)	(0.012)	(0.002)	(0.003)	(0.001)	(0.001)	(0.006)	(0.006)
BP<140	−0.002	0.020***	0.004***	0.005***	−0.039***	−0.003***	0.059***	0.007*
	(0.004)	(0.004)	(0.001)	(0.002)	(0.001)	(0.0003)	(0.003)	(0.004)
Observations	619,907	619,907	619,907	619,907	619,907	619,907	619,907	619,907
Mean dep. var.	0.566	0.594	0.96	0.937	4.318	5.065	0.641	0.286
Clinics	257	257	257	257	257	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences results estimated using equation (2) on placebo outcomes. BP< 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of bunching at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. DM2 stands for type 2 diabetes. Tests done at encounter and new diagnosis variables are indicators. Weight is measured in kilograms and log transformed. Height is measured in centimeters and log transformed. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for male, 1-year age, year and quarter of primary care visit.

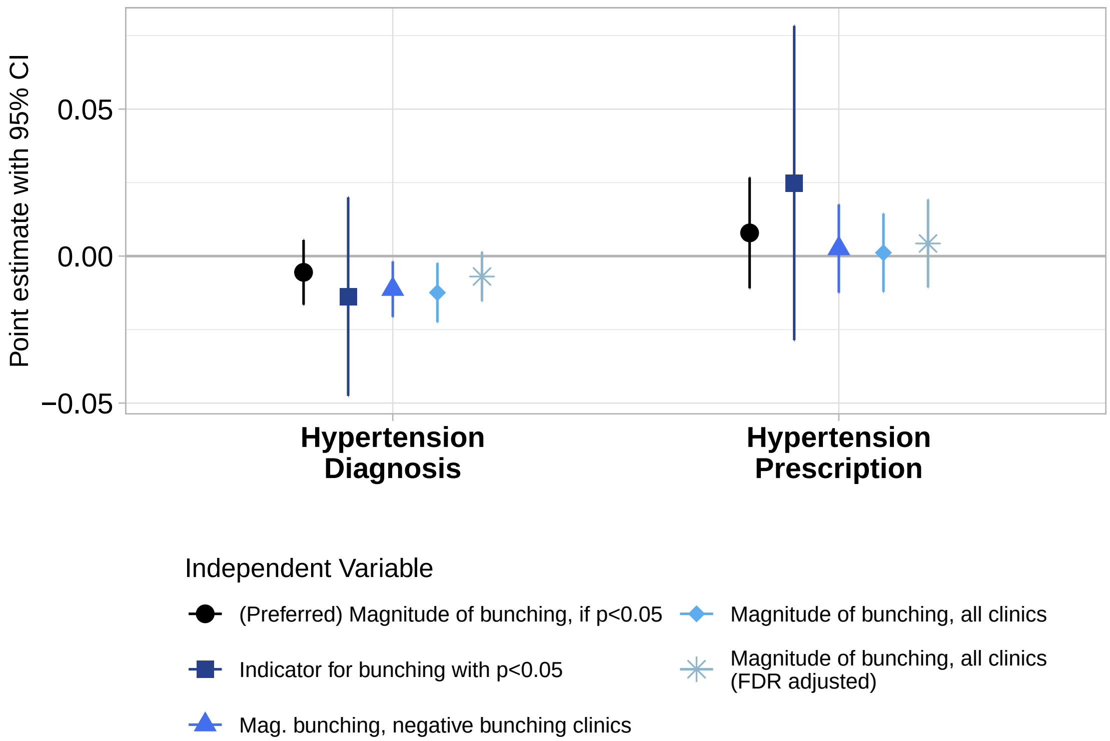
A.9 Robustness: Parameterization of Discretion

For sensitivity analyses I construct four other variables that measure discretion. First, the magnitude of bunching for all clinics (regardless of statistical significance), normalized to mean zero, standard deviation one. Second, the magnitude of bunching for clinics with negative bunching, regardless of statistical significance, normalized to mean zero, standard deviation one. Third, I examine the extensive margin by constructing an indicator for if the clinic's bunching estimate is negative and statistically significant, zero otherwise. Last, I apply an adaptive shrinkage estimator [Stephens et al., 2016], to adjust for the false discovery rate (FDR) in the dependent variable, which represents the magnitude of bunching across all clinics, regardless of statistical significance. I then normalize this variable to have a mean of zero and a standard deviation of one.

In each of the independent variables constructed, a higher value indicates more excess mass below the diagnostic threshold and missing mass above it. The four independent variables yield very similar estimated effects. Figures A6, A7, and A8 display coefficients and 95% confidence intervals for all main outcomes estimated with each of the four independent variables.

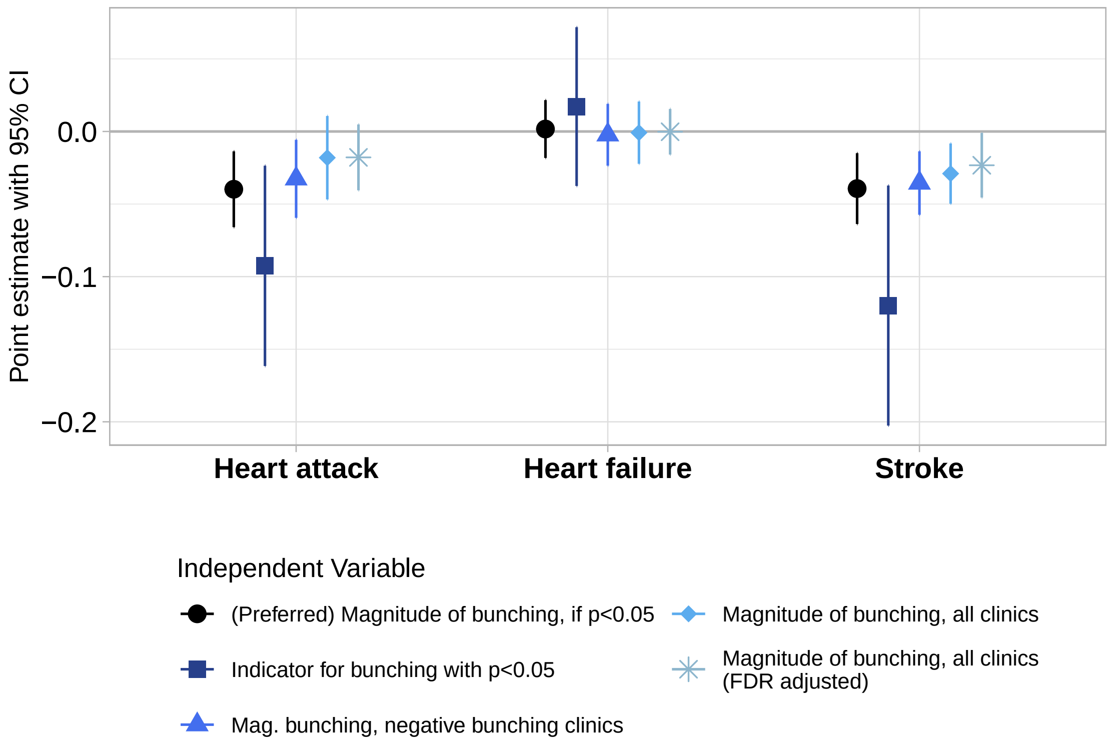
For the FDR-adjusted version of the independent variable, I also present full versions of the main results tables (A10 to A12). Here, while the lower rate of heart attack among patients exposed to discretion is no longer statistically significant, the results are qualitatively the same: discretion does not impact adherence to clinical guidelines, but patients affected by discretion are of lower cardiovascular risk (measured by strokes), and are more likely to be older males.

Figure A6: Comparison of results under different parameterizations of the magnitude of discretion: adherence to clinical practice guidelines



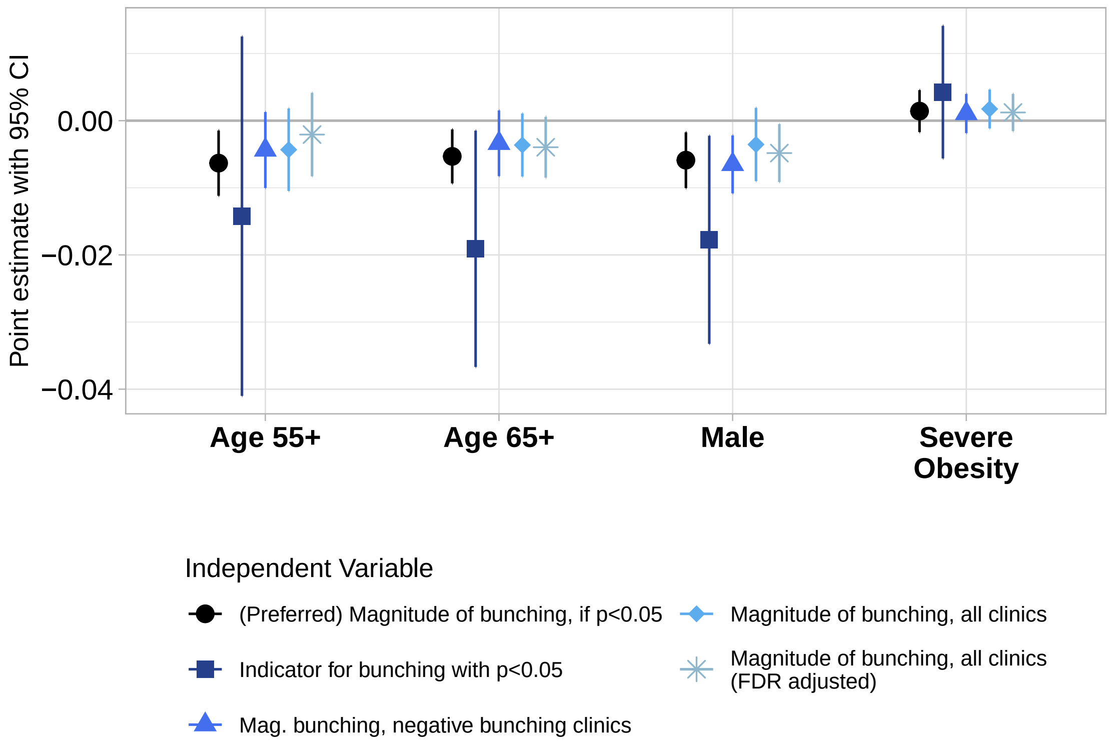
Note: This figure presents the impact of discretion on key outcomes, under different discretion specifications. Point estimates for (Discretion x BP<140) and 95% confidence intervals shown from equation (2) are shown. Each outcome is estimated five separate times with different treatment variables. The preferred treatment variable, used everywhere else in this paper, is shown in black. This is the magnitude of discretion among clinics with negative discretion that was different from zero at alpha=0.05 significance level, zero otherwise. The treatment variable with the square marker is an indicator for if the clinic had a magnitude of negative discretion that was different from zero at alpha=0.05 significance level, zero otherwise. The treatment variable with the triangle marker is the magnitude of discretion for clinics with negative discretion (regardless of statistical significance), normalized to mean zero, SD 1. The treatment variable with the diamond marker is the magnitude of discretion for all clinics (regardless of statistical significance), normalized to mean zero, SD 1. The treatment variable with the star marker is the magnitude of discretion for all clinics (regardless of statistical significance), false discovery rate (FDR) adjusted and normalized to mean zero, SD 1. Diagnosis and prescription are indicators for if the patient received a new diagnosis or prescription at the visit.

Figure A7: Comparison of results under different parameterizations of the magnitude of discretion: cardiovascular hospitalizations



Note: This figure presents the impact of discretion on key outcomes, under different discretion specifications. Point estimates for (Discretion x BP<140) and 95% confidence intervals shown from equation (2) are shown. Each outcome is estimated five separate times with different treatment variables. The preferred treatment variable, used everywhere else in this paper, is shown in black. This is the magnitude of discretion among clinics with negative discretion that was different from zero at alpha=0.05 significance level, zero otherwise. The treatment variable with the square marker is an indicator for if the clinic had a magnitude of negative discretion that was different from zero at alpha=0.05 significance level, zero otherwise. The treatment variable with the triangle marker is the magnitude of discretion for clinics with negative discretion (regardless of statistical significance), normalized to mean zero, SD 1. The treatment variable with the diamond marker is the magnitude of discretion for all clinics (regardless of statistical significance), normalized to mean zero, SD 1. The treatment variable with the star marker is the magnitude of discretion for all clinics (regardless of statistical significance), false discovery rate (FDR) adjusted and normalized to mean zero, SD 1. Stroke, heart attack, and heart failure each are indicators for hospitalization with the event within six months of the primary care visit, and are each multiplied by 100 so they can be interpreted as percentages.

Figure A8: Comparison of results under different parameterizations of the magnitude of discretion: patient characteristics



Note: This figure presents the impact of discretion on key outcomes, under different discretion specifications. Point estimates for (Discretion x BP<140) and 95% confidence intervals shown from equation (2) are shown. Each outcome is estimated five separate times with different treatment variables. The preferred treatment variable, used everywhere else in this paper, is shown in black. This is the magnitude of discretion among clinics with negative discretion that was different from zero at alpha=0.05 significance level, zero otherwise. The treatment variable with the square marker is an indicator for if the clinic had a magnitude of negative discretion that was different from zero at alpha=0.05 significance level, zero otherwise. The treatment variable with the triangle marker is the magnitude of discretion for clinics with negative discretion (regardless of statistical significance), normalized to mean zero, SD 1. The treatment variable with the diamond marker is the magnitude of discretion for all clinics (regardless of statistical significance), normalized to mean zero, SD 1. The treatment variable with the star marker is the magnitude of discretion for all clinics (regardless of statistical significance), false discovery rate (FDR) adjusted and normalized to mean zero, SD 1. Each dependent variable is an indicator for if the patient has the characteristic. Severe obesity is body mass index ≥ 35.

Table A10: Impact of Discretion on Hypertension Diagnosis and Prescription (FDR adjusted)

	Hypertension Diagnosis	Hypertension Prescription
	(1)	(2)
Discretion x BP<140	-0.007	0.004
	(0.004)	(0.008)
Discretion	-0.009	-0.023^{*}
	(0.007)	(0.013)
BP<140	-0.251^{***}	-0.169^{***}
	(0.005)	(0.005)
Observations	619,907	619,907
Mean dep. var.	0.522	0.376
Clinics	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences results estimated using equation (2). BP< 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of bunching at the clinic (normalized to mean 0 SD 1, among all clinics, false discovery rate adjusted). The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for year and quarter of primary care visit. Hypertension diagnosis is an indicator for if the patient was newly diagnosed. Hypertension prescription is an indicator for if the patient was prescribed a new hypertension medication.

Table A11: Impact of Discretion on Cardiovascular Hospitalization 3, 6, and 12 Months After Primary Care Visit (FDR adjusted)

	Stroke			Heart Attack			Congestive Heart Failure		
	$\leq$3 mo.	$\leq$6 mo.	$\leq$12 mo.	$\leq$3 mo.	$\leq$6 mo.	$\leq$12 mo.	$\leq$3 mo.	$\leq$6 mo.	$\leq$12 mo.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Discretion x BP<140	-0.012^{*}	-0.023^{**}	-0.042^{**}	-0.009	-0.018	-0.015	0.005	-0.0002	-0.014
	(0.007)	(0.011)	(0.017)	(0.008)	(0.012)	(0.021)	(0.006)	(0.008)	(0.013)
Discretion	0.016^{**}	0.027^{**}	0.048^{***}	0.004	0.009	0.004	-0.003	-0.003	0.003
	(0.007)	(0.011)	(0.016)	(0.008)	(0.011)	(0.019)	(0.004)	(0.006)	(0.011)
BP<140	-0.029^{***}	-0.061^{***}	-0.106^{***}	-0.025^{***}	-0.062^{***}	-0.119^{***}	0.010	0.013	0.004
	(0.005)	(0.009)	(0.014)	(0.009)	(0.013)	(0.019)	(0.007)	(0.010)	(0.014)
Observations	619,907	619,907	619,907	619,907	619,907	619,907	619,907	619,907	619,907
Mean dep. var.(%)	0.029	0.066	0.137	0.109	0.205	0.385	0.062	0.124	0.232
Clinics	257	257	257	257	257	257	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences results estimated using equation (2). BP< 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of bunching at the clinic (normalized to mean 0 SD 1, among all clinics, false discovery rate adjusted). The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. All dependent variables in this table are multiplied by 100. Each dependent variable is an indicator for if the patient was hospitalized with the listed condition, within 3, 6 or 12 months of their primary care visit, zero otherwise. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for male, 1-year age, year and quarter of primary care visit. Stroke includes any cerebral infarction (ICD-10 codes I63). Heart attack includes acute coronary syndrome, myocardial infarction, and any other condition associated with sudden, reduced blood flow to the heart (I20-22, I24-25). Congestive heart failure is I50.

Table A12: Selection by Patient Characteristics Representative of High Cardiovascular Risk (FDR adjusted)

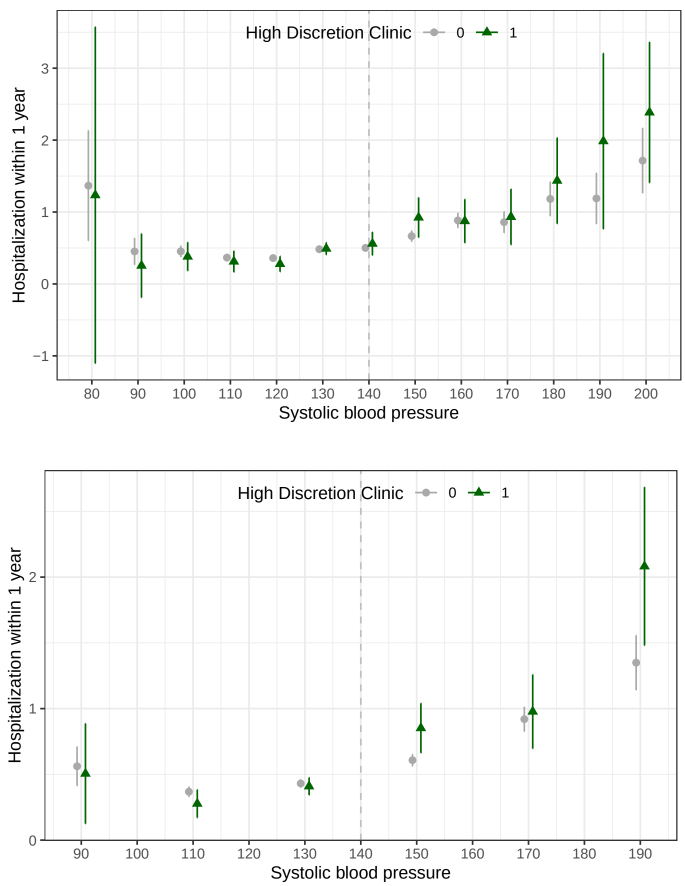
	Male	Severe Obesity	Age 55+	Age 65+
	(1)	(2)	(3)	(4)
Discretion x BP<140	-0.005^{**}	0.001	-0.002	-0.004^{*}
	(0.002)	(0.001)	(0.003)	(0.002)
Discretion	0.005^{***}	-0.004	-0.005	-0.007
	(0.002)	(0.003)	(0.005)	(0.005)
BP<140	-0.096^{***}	-0.019^{***}	-0.141^{***}	-0.128^{***}
	(0.002)	(0.001)	(0.004)	(0.003)
Observations	619,907	582,880	619,907	619,907
Mean dep. var.	0.382	0.067	0.596	0.336
Clinics	257	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences results estimated using equation (2). BP< 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of bunching at the clinic (normalized to mean 0 SD 1, among all clinics, false discovery rate adjusted). The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Each dependent variable is an indicator for if the patient has the characteristic listed, zero otherwise. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for year and quarter of primary care visit.

A.10 Robustness: Narrowing the Window Around the Diagnostic Threshold

Bunching estimation finds that healthcare providers manipulate blood pressure, so my preferred analyses do not conditioning on blood pressure (beyond above vs. below 140), as this may bias the results. However, showing that the impact of discretion on hospitalizations is driven by systolic blood pressure values were we expect missing or excess mass to occur can help support the finding that provider discretion leads to better sorting by patient risk. In figure A9 I plot mean hospitalization with stroke or heart attack, by 10 and 20 unit bins of systolic blood pressure, and stratified by clinics with vs. without bunching significant at the 5% level. Bars indicate 95% confidence intervals and are adjusted for clustering at the clinic level. Second, I also re-estimate main results tables (tables 2, 3, and 4) conditional on patients with a recorded systolic blood pressure of 90-160 mmHg, and 120-160 mmHg.

Figure A9: 1-year Hospitalization with Stroke or Heart Attack, by Systolic Blood Pressure



Note: Mean 1-year hospitalization for stroke or heart attack, by 10 or 20 unit bins of systolic blood pressure, and stratified by high vs. low discretion clinic. Bars indicate 95% confidence intervals and are adjusted for clustering at the clinic level.

Table A13: Impact of Discretion on Hypertension Diagnosis and Prescription (BP=90 to 160)

	Hypertension Diagnosis	Hypertension Prescription
	(1)	(2)
Discretion x BP<140	-0.004	0.008
	(0.006)	(0.009)
Discretion	-0.011^{*}	-0.021
	(0.006)	(0.014)
BP<140	-0.239^{***}	-0.160^{***}
	(0.005)	(0.005)
Observations	576,201	576,201
Mean dep. var.	0.522	0.376
Clinics	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table replicates table 2 among patients with systolic blood pressure 90-160 mmHg.

Table A14: Impact of Discretion on Hypertension Diagnosis and Prescription (BP=120 to 160)

	Hypertension Diagnosis	Hypertension Prescription
	(1)	(2)
Discretion x BP<140	-0.004	0.006
	(0.005)	(0.007)
Discretion	-0.010^{*}	-0.020
	(0.006)	(0.013)
BP<140	-0.185^{***}	-0.125^{***}
	(0.004)	(0.004)
Observations	434,581	434,581
Mean dep. var.	0.522	0.376
Clinics	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table replicates table 2 among patients with systolic blood pressure 120-160 mmHg.

Table A15: Impact of Discretion on Cardiovascular Hospitalization 3, 6, and 12 Months After Primary Care Visit (BP=90 to 160)

	≤3 mo.	≤6 mo.	≤12 mo.	≤3 mo.	≤6 mo.	≤12 mo.	≤3 mo.	≤6 mo.	≤12 mo.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Discretion x BP<140	−0.010	−0.031***	−0.046***	−0.038***	−0.045***	−0.042*	−0.001	−0.001	−0.007
	(0.007)	(0.010)	(0.017)	(0.015)	(0.015)	(0.022)	(0.007)	(0.008)	(0.013)
Discretion	0.009	0.031***	0.045***	0.032**	0.038**	0.037*	−0.001	−0.005	−0.004
	(0.007)	(0.010)	(0.015)	(0.014)	(0.016)	(0.022)	(0.006)	(0.007)	(0.011)
BP<140	−0.022***	−0.048***	−0.072***	−0.015	−0.044***	−0.075***	0.018**	0.029***	0.032**
	(0.006)	(0.009)	(0.014)	(0.010)	(0.014)	(0.020)	(0.007)	(0.011)	(0.015)
Observations	576,201	576,201	576,201	576,201	576,201	576,201	576,201	576,201	576,201
Mean dep. var.(%)	0.029	0.066	0.137	0.109	0.205	0.385	0.062	0.124	0.232
Clinics	257	257	257	257	257	257	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table replicates table 3 among patients with systolic blood pressure 90-160 mmHg.

Table A16: Impact of Discretion on Cardiovascular Hospitalization 3, 6, and 12 Months After Primary Care Visit (BP=120 to 160)

	≤3 mo.	≤6 mo.	≤12 mo.	≤3 mo.	≤6 mo.	≤12 mo.	≤3 mo.	≤6 mo.	≤12 mo.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Discretion x BP<140	−0.011*	−0.031***	−0.042**	−0.037**	−0.043**	−0.040*	−0.001	0.002	−0.004
	(0.007)	(0.011)	(0.018)	(0.016)	(0.017)	(0.024)	(0.008)	(0.010)	(0.015)
Discretion	0.009	0.031***	0.044***	0.032**	0.038**	0.036*	−0.001	−0.006	−0.004
	(0.007)	(0.010)	(0.015)	(0.014)	(0.016)	(0.022)	(0.006)	(0.007)	(0.011)
BP<140	−0.022***	−0.047***	−0.077***	−0.025**	−0.050***	−0.078***	0.004	0.005	−0.006
	(0.006)	(0.010)	(0.014)	(0.011)	(0.015)	(0.021)	(0.008)	(0.011)	(0.016)
Observations	434,581	434,581	434,581	434,581	434,581	434,581	434,581	434,581	434,581
Mean dep. var.(%)	0.029	0.066	0.137	0.109	0.205	0.385	0.062	0.124	0.232
Clinics	257	257	257	257	257	257	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table replicates table 3 among patients with systolic blood pressure 120-160 mmHg.

Table A17: Selection by Patient Characteristics Representative of High Cardiovascular Risk (BP=90 to 160)

	Male	Severe Obesity	Age 55+	Age 65+
	(1)	(2)	(3)	(4)
Discretion x BP<140	−0.005**	0.001	−0.003	−0.002
	(0.002)	(0.002)	(0.003)	(0.002)
Discretion	0.007***	0.0003	−0.009*	−0.012**
	(0.002)	(0.003)	(0.005)	(0.005)
BP<140	−0.089***	−0.018***	−0.123***	−0.110***
	(0.002)	(0.001)	(0.003)	(0.003)
Observations	576,201	542,295	576,201	576,201
Mean dep. var.	0.382	0.067	0.596	0.336
Clinics	257	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table replicates table 4 among patients with systolic blood pressure 90-160 mmHg.

Table A18: Selection by Patient Characteristics Representative of High Cardiovascular Risk (BP=120 to 160)

	Male	Severe Obesity	Age 55+	Age 65+
	(1)	(2)	(3)	(4)
Discretion x BP<140	−0.003	0.001	−0.003	−0.002
	(0.003)	(0.001)	(0.002)	(0.002)
Discretion	0.007***	0.0003	−0.009*	−0.012**
	(0.002)	(0.003)	(0.005)	(0.005)
BP<140	−0.066***	−0.012***	−0.094***	−0.087***
	(0.002)	(0.001)	(0.003)	(0.003)
Observations	434,581	408,880	434,581	434,581
Mean dep. var.	0.382	0.067	0.596	0.336
Clinics	257	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table replicates table 4 among patients with systolic blood pressure 120-160 mmHg.

A.11 Impact of Discretion among Alive Patients

To definitively conclude that discretion leads to a more accurate sorting of patients on either side of the diagnostic threshold in terms of health I would ideally measure effects on patient mortality. One limitation of this analysis is that mortality is unobserved. To address this limitation I re-estimate the effect of discretion on cardiovascular hospitalizations among patients who are known to be alive for at least 3, 6, or 12 months after their initial primary care visit. For this group, admission to the hospital for a cardiovascular event is considered the most severe adverse health outcome related to high blood pressure.

In practice, if a patient has another primary care visit or is hospitalized after the initial visit, we can conclude they were alive during the interim. 85.4% of patients are re-observed 3 months after their first visit, 82.7% within 6 months, and 77.0% within 12 months (table A19). Patients who are not observed during this time period may have left the country, transitioned to private health care, simply not had a scheduled appointment nor fallen ill, or may have died. However, we cannot distinguish between these events.

First, I find that retention in the sample is unrelated to clinic level discretion (Table A20), and so in Tables A21 and A22 each model is estimated conditional on the patient being re-observed during the relevant time frame.

I find no impact of discretion on all-cause hospitalization among the retained sample (Table A21). After conditioning on retained patients, the effect of discretion on cardiovascular hospitalization is similar to when the full sample is used: a one standard deviation in discretion is associated with a lower rate of stroke hospitalization of 0.02 to 0.06 within 3 to 6 months. For heart attack the results are also similar: a one standard deviation in discretion is associated with a lower rate of stroke hospitalization of 0.02 to 0.03 within 3 to 6 months. While the magnitudes are very similar to when the full sample is used (see Table 3), the estimates are slightly less precise, likely driven by the smaller number of observations used here. Overall, these results confirm that discretion leads to improved sorting of patients by risk with respect to the diagnostic threshold.

Table A19: Sample of Primary Care 3, 6, and 12 Months After Initial Visit

	Yes	No	% Yes	% No
Observed again after 3 months	529,497	90,410	85.4	14.6
Observed again after 6 months	512,475	107,432	82.7	17.3
Observed again after 12 months	477,079	142,828	77.0	23.0

Note: This table describes the number of patients who were observed a second time (or more) in the electronic health records or hospitalization data, 3, 6, or 12 months after their initial primary care visit. Because they were observed again these patients are considered living during this period.

Table A20: Impact of Bunching on Time Patient is Observed in the Data

	Retained 3 mo.	**Retained 6 mo.**	**Retained 12 mo.**
	(1)	(2)	(3)
Bunch x BP<140	0.003	0.002	0.002
	(0.003)	(0.003)	(0.003)
Bunch	-0.003	-0.005	-0.007
	(0.005)	(0.005)	(0.007)
BP<140	-0.006^{***}	-0.006^{***}	-0.005^{**}
	(0.002)	(0.002)	(0.002)
Observations	619,907	619,907	619,907
Mean dep. var.	0.854	0.827	0.77
Clinics	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences results estimated using equation (2), of the impact of discretion on the probability that the patient was re-observed in the data within 3, 6, or 12 months. BP< 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of bunching at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for male, 1-year age, year and quarter of primary care visit.

Table A21: Impact of Discretion on Any Hospitalization 3, 6, and 12 Months After Primary Care Visit, Among Retained Patients

	Hosp <3 months	Hosp <6 months	Hosp <12 months
	(1)	(2)	(3)
Discretion x BP<140	−0.072	−0.089	−0.133
	(0.055)	(0.071)	(0.108)
Discretion	0.038	0.062	0.147
	(0.060)	(0.088)	(0.124)
BP<140	−0.014	−0.233***	−0.498***
	(0.051)	(0.075)	(0.098)
Observations	529,497	512,475	477,079
Mean dep. var.(%)	2.419	4.621	8.508
Clinics	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences results estimated using equation (2). BP< 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of bunching at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for male, 1-year age, year and quarter of primary care visit. Model 1 is estimated among patients who returned to primary care or were hospitalized 3 or more months after their initial primary care visit (85.4% of the full sample). Model 2; 6 or more months (82.7% of the full sample). Model 3; 12 or more months (77.0% of the full sample). Dependent variables are an indicator for hospitalization for any reason within 3, 6, or 12 months of the patient's initial primary care visit.

Table A22: Impact of Discretion on Cardiovascular Hospitalization 3, 6, and 12 Months After Primary Care Visit, Among Retained Patients

	Within 3 months			Within 6 months			Within 12 months		
	Stroke	**Heart attack**	**Heart failure**	**Stroke**	**Heart attack**	**Heart failure**	**Stroke**	**Heart attack**	**Heart failure**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Discretion x BP<140	-0.017	-0.020^{**}	0.007	-0.036^{***}	-0.024^{*}	0.007	-0.057^{***}	-0.025	-0.0001
	(0.011)	(0.009)	(0.006)	(0.014)	(0.013)	(0.011)	(0.021)	(0.024)	(0.015)
Discretion	0.018^{*}	0.013	-0.007	0.038^{***}	0.021	-0.010	0.057^{***}	0.020	-0.011
	(0.011)	(0.008)	(0.004)	(0.014)	(0.014)	(0.009)	(0.021)	(0.022)	(0.010)
BP<140	-0.027^{***}	-0.019^{**}	0.007	-0.064^{***}	-0.055^{***}	0.0002	-0.103^{***}	-0.116^{***}	-0.026^{*}
	(0.006)	(0.010)	(0.007)	(0.010)	(0.014)	(0.011)	(0.015)	(0.021)	(0.016)
Observations	529,497	529,497	529,497	512,475	512,475	512,475	477,079	477,079	477,079
Mean dep. var.(%)	0.027	0.106	0.055	0.063	0.198	0.109	0.124	0.368	0.195
Clinics	257	257	257	257	257	257	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences results estimated using equation (2). BP< 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of bunching at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for male, 1-year age, year and quarter of primary care visit. Model 1 is estimated among patients who returned to primary care or were hospitalized 3 or more months after their initial primary care visit (85.4% of the full sample). Model 2; 6 or more months (82.7% of the full sample). Model 3; 12 or more months (77.0% of the full sample). Dependent variables are an indicator for hospitalization for any reason within 3, 6, or 12 months of the patient's initial primary care visit. Stroke includes any cerebral infarction (ICD-10 codes I63). Heart attack includes acute coronary syndrome, myocardial infarction, and any other condition associated with sudden, reduced blood flow to the heart (I20-22, I24-25). Congestive heart failure (CHF) is I50.

A.12 Other Possible Heuristic Variables

Table A23: Patient Characteristics - Age

	Age 50+	Age 60+	Age 70+
	(1)	(2)	(3)
Discretion x BP<140	−0.0002	−0.003	−0.003
	(0.002)	(0.002)	(0.002)
Discretion	0.006**	−0.002	−0.007*
	(0.003)	(0.002)	(0.004)
BP<140	0.033***	−0.063***	−0.086***
	(0.002)	(0.002)	(0.002)
Observations	619,907	619,907	619,907
Mean dep. var.	0.263	0.264	0.201
Clinics	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences results estimated using equation (2). BP< 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of discretion at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Each dependent variable is an indicator for if the patient is at least 50, 60, or 70 years old, zero otherwise. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for year and quarter of primary care visit.

Table A24: Selection by Patient Characteristics - Health History Questions

	Was asked about family history of:				Patient has family history of:			
	Any	Heart disease	Dyslipidemia	CVD	Any	Heart disease	Dyslipidemia	CVD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Discretion x BP<140	−0.006	−0.005	−0.004	−0.004	−0.002	−0.0004	0.0001	−0.001
	(0.005)	(0.004)	(0.005)	(0.004)	(0.002)	(0.002)	(0.0004)	(0.001)
Discretion	−0.001	−0.0004	−0.004	−0.002	−0.007	−0.007	−0.001	−0.002
	(0.012)	(0.012)	(0.013)	(0.012)	(0.005)	(0.004)	(0.001)	(0.002)
BP<140	−0.022***	−0.022***	−0.021***	−0.021***	−0.012***	−0.008***	0.004***	−0.009***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.002)	(0.002)	(0.001)	(0.001)
Observations	619,907	619,907	619,907	619,907	360,053	378,879	359,547	372,205
Mean dep. var.	0.63	0.611	0.58	0.6	0.212	0.139	0.024	0.067
Clinics	257	257	257	257	257	257	257	257

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents difference-in-differences results estimated using equation (2). BP< 140 is an indicator for if recorded systolic blood pressure (BP) was below 140. Discretion is a continuous variable for the estimated magnitude of bunching at the clinic, normalized to mean 0, standard deviation 1. The coefficient of interest is the interaction term, which is interpreted as the impact of discretionary clinical decision-making. Each dependent variable is an indicator for if the patient has the characteristic listed, zero otherwise. Robust standard errors clustered at the primary care clinic level are in parentheses. All models include fixed effects for year and quarter of primary care visit. Models 1-4 are indicators for if any answer to the health history question was recorded in the EHR. Models 5-8 are estimated among patients who did have the relevant health history question recorded in the EHR. CVD stands for cardiovascular disease.

A.13 Provider Private Information

In addition to using heuristics to help make diagnostic decisions, providers at high discretion clinics may be using private information. Here, private information refers to anything that is observable to the provider, but not recorded in the electronic medical record. Examples include some patient characteristics and medical history, or the many factors that influence the quality of blood pressure measurement, such as whether the patient was talking or had their legs crossed during measurement (see Table A4 for Chile's guidelines on factors that may reduce the quality of blood pressure measurement).

To test for the importance of private information I again rely on the fact that the primary goal of treating hypertension is to reduce the patient's risk of future adverse cardiac events, like stroke or heart attack. I first measure the variation in cardiovascular hospitalizations explained by systolic blood pressure alone, or systolic blood pressure with observable patient characteristics.

Specifically, I regress an indicator for stroke and/or heart attack in the year following a primary care visit onto blood pressure alone, and onto blood pressure plus a vector of patient characteristics using the following equation for results in tables A25 and A26:

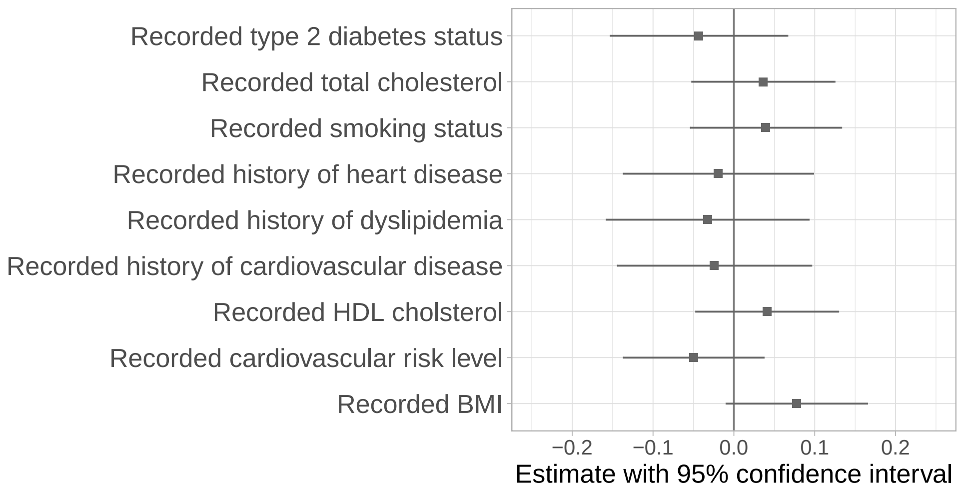
$$Y_{i,c} = \alpha + \beta_1 BP_i + \beta_2 X_i + \varepsilon_{ic} \tag{3}$$

Where $Y_{i,c}$ is an indicator for if patient i at clinic c was hospitalized for stroke or heart attack (ACS) within 365 days of their primary care visit. BP_i is systolic blood pressure (continuous in table A25, or an indicator for if SBP was recorded as 140 or higher table A26). X_i are patient characteristics from the Framingham 10-year cardiovascular risk score [Anderson et al., 1991]: an indicator for previous diagnosis with type 2 diabetes (vs. no diagnosis), an indicator for male (vs. female), log age at visit, log total cholesterol (mg/dL), log HDL cholesterol (mg/dL), and an indicator for self reported smoker (vs. non-smoker). Standard errors were clustered at the clinic level. Models are estimated separately among clinics with a magnitude of bunching greater than zero (discretion clinics), compared to clinics with zero bunching (non-discretion clinics; both at $p<0.05$). Results are similar if only stroke, only ACS, are used, or if 180 or 90 day windows are used. I test both continuous blood pressure (table A25) or blood pressure classified into above vs. below the diagnostic threshold of 140 mmHg (table A26).

Among patients at non-discretion clinics, systolic blood pressure explains 0.1% of the variation in future events (adjusted R-squared, column 1 table A25). After adding in relevant, observable patient characteristics, 0.3% of the variation is explained. The explanatory power of blood pressure is approximately two times higher at discretion clinics, where blood pressure may include both signal about the patient's health, and signal about the provider's opinion about the patient's health (adjusted R-squared = 0.002, column 1 table A25). Still, more than 99% of the variation is unexplained after accounting for clinically relevant information that is observable in the electronic health record. Because we have seen that providers are able to make relatively good predictions about future risk and reclassify patients accordingly (section 5.3), these findings suggest they may also be using of private information not accounted for in these models. Finally, the increase in adjusted R-squared at discretion clinics compared to non-discretion clinics suggests more incorporation of private information into the physician's decision at those clinics.

It is possible the data quality varies across clinics, given that administrative data like electronic health records rely on providers documenting information during or after a patient's visit. Here I test for whether the presence of variables listed in the Framingham cardiovascular risk score, or health history questions, is differential by the estimated magnitude of clinic discretion. Specifically, I regress the magnitude of discretion at the clinic level on an indicator for if the patient health or history was recorded in the EHR during the visit and zero otherwise, plus year and quarter fixed effects, with standard errors clustered at the clinic level. Figure A10 displays point estimates and 95% confidence intervals from these univariate regressions, and we observe no systematic difference in the quality of EHR data across clinic discretion. So, for this exercise missing values were imputed with the population mean.

Figure A10: Balance in the existence of patient cardiovascular and history information



Note: This figure tests for if patient health and history is recorded differently by magnitude of discretion. Point estimates and 95% confidence intervals shown from the regression of clinics' continuous and standardized bunching magnitude on indicators for if a patient health measure or history question was non-missing. Each variable is from a separate regression and includes year and quarter fixed effects, with standard errors clustered at the clinic level.

Table A25: Predicting Stroke and Heart Attack with Patient Characteristics

	Stroke or Heart Attack ≤12 Months (per 100)			
	(1)	(2)	(3)	(4)
Blood Pressure	0.009***	0.006***	0.016***	0.012***
	(0.001)	(0.001)	(0.002)	(0.002)
DM2 diagnosis		0.076***		0.160*
		(0.024)		(0.095)
Male		0.440***		0.362***
		(0.025)		(0.073)
Log Age		0.903***		1.079***
		(0.038)		(0.142)
Log Total Chol.		0.245***		−0.543
		(0.089)		(0.830)
Log HDL Chol.		−0.241***		−0.273
		(0.041)		(0.168)
Smoker		0.238***		0.357
		(0.068)		(0.276)
Constant	−0.681***	−4.714***	−1.541***	−1.168
	(0.082)	(0.567)	(0.294)	(4.764)
Observations	560,220	560,220	59,687	59,687
Adjusted R^2	0.001	0.003	0.002	0.004
Mean dep. var.(%)	0.514	0.514	0.576	0.576
Clinics	233	233	24	24

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents OLS estimates of the association between blood pressure and other patient characteristics and the 1-year risk of hospitalization with stroke or heart attack. Blood pressure is continuous systolic. DM2 stands for type 2 diabetes. Chol. stands for cholesterol. Models 1-2 were estimated among patients at clinics with no statistically significant bunching. Models 3-4 were were estimated among patients at clinics with statistically significant bunching (p<0.05). Standard errors were clustered at the clinic level.

Table A26: Predicting Stroke and Heart Attack with Patient Characteristics

	Stroke or Heart Attack ≤12 Months (per 100)			
	(1)	(2)	(3)	(4)
Blood Pressure≥140	0.320***	0.201***	0.620***	0.471***
	(0.024)	(0.024)	(0.103)	(0.098)
DM2 diagnosis		0.077***		0.167*
		(0.024)		(0.095)
Male		0.446***		0.375***
		(0.025)		(0.072)
Log Age		0.927***		1.119***
		(0.038)		(0.138)
Log Total Chol.		0.246***		−0.541
		(0.089)		(0.829)
Log HDL Chol.		−0.242***		−0.276
		(0.041)		(0.168)
Smoker		0.237***		0.355
		(0.068)		(0.275)
Constant	0.414***	−4.118***	0.371***	0.086
	(0.014)	(0.554)	(0.031)	(4.761)
Observations	560,220	560,220	59,687	59,687
Adjusted R^2	0.0004	0.003	0.001	0.004
Mean dep. var.(%)	0.514	0.514	0.576	0.576
Clinics	233	233	24	24

Note: *p<0.1; **p<0.05; ***p<0.01. This table presents OLS estimates of the association between blood pressure and other patient characteristics and the 1-year risk of hospitalization with stroke or heart attack. Blood pressure ≥ 140 is an indicator for if recorded systolic blood pressure (BP) was at or above 140. DM2 stands for type 2 diabetes. Chol. stands for cholesterol. Models 1-2 were estimated among patients at clinics with no statistically significant bunching. Models 3-4 were were estimated among patients at clinics with statistically significant bunching (p<0.05). Standard errors were clustered at the clinic level.

A.14 Subsequent Visits

Table A27: Clinical Actions at Subsequent Visits

Group	**Diagnosed at first visit**		**Diagnosed at a follow up visit**		**Never diagnosed**	
Significant negative bunching	**0**	**1**	**0**	**1**	**0**	**1**
N patients	366,960 (65.5%)	36,333 (60.9%)	31,074 (5.5%)	4,342 (7.3%)	162,162 (28.9%)	19,010 (31.8%)
N visits	366,960 (25.2%)	36,333 (10.1%)	146,877 (64.2%)	19,132 (1.8%)	938,579 (88.6%)	101,772 (9.6%)
Mean follow up visits until diagnosis	-	-	2.4	2.5	-	-
Mean days until diagnosis	-	-	783.1	715.3	-	-
Share of patients diagnosed at their second visit	-	-	49.9%	46.6%	-	-

Note: Patients grouped into those who were diagnosed at their first visit, those diagnosed at a subsequent visit, and those never diagnosed during the study period. They are also grouped into those with statistically significant negative bunching at their clinic, vs. not. Percent of total visits or encounters are shown in parentheses, out of a total of 560,196 patients and 1,452,416 visits at clinics without statistically significant negative bunching, and 59,685 patients and 1,059,483 visits at clinics with statistically significant negative bunching. In bunching analyses, one visit per patient is used. Here, all visits up to and including the diagnostic visit (if it occurs) are included. Notably, this sample is selected with respect to bunching.